Mol Genet Genomics (2008) 280:187–198
DOI 10.1007/s00438-008-0355-0

ORIGINAL PAPER

Recent duplications dominate NBS-encoding gene expansion in two woody species

Sihai Yang · Xiaohui Zhang · Jia-Xing Yue · Dacheng Tian · Jian-Qun Chen

Received: 14 January 2008 / Accepted: 1 June 2008 / Published online: 19 June 2008

Abstract Most disease resistance genes in plants encode NBS-LRR proteins. However, in woody species, little is known about the evolutionary history of these genes. Here, we identified 459 and 330 respective NBS-LRRs in grapevine and poplar genomes. We subsequently investigated protein motif composition, phylogenetic relationships and physical locations. We found significant excesses of recent duplications in perennial species, compared with those of annuals, represented by rice and *Arabidopsis*. Consequently, we observed higher nucleotide identity among paralogs and a higher percentage of NBS-encoding genes positioned in numerous clusters in the grapevine and poplar. These results suggested that recent tandem duplication played a major role in NBS-encoding gene expansion in perennial species. These duplication events, together with a higher probability of recombination revealed in this study, could compensate for the longer generation time in woody perennial species e.g. duplication and recombination could serve to generate novel resistance specificities. In addition, we observed extensive species-specific expansion in TIR-NBS-encoding genes. Non-TIR-NBS-encoding genes were poly- or paraphyletic, i.e. genes from three or more plant species were nested in different clades, suggesting different evolutionary patterns between these two gene types.

Keywords NBS-LRR genes · Disease resistance genes · Evolution · Grapevine · Poplar

Sihai Yang and Xiaohui Zhang contributed equally to this work.

Communicated by A. Tyagi.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-008-0355-0) contains supplementary material, which is available to authorized users.

S. Yang · X. Zhang · J.-X. Yue · D. Tian (✉) · J.-Q. Chen (✉)
State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, 210093 Nanjing, China
e-mail: dtian@nju.edu.cn

J.-Q. Chen
e-mail: chenjq@nju.edu.cn

Introduction

Numerous disease resistance genes (*R*-genes) conferring resistance to a diverse spectrum of pathogens, including bacteria, fungi, oomycetes, viruses, and nematodes, have been isolated from a wide range of plant species (Dangl and Jones 2001; Meyers et al. 2003; McHale et al. 2006). The NBS-LRR gene, which contain a nucleotide-binding site (NBS) and leucine-rich repeats (LRRs), is the largest class of *R*-genes. These proteins have been shown to function as intracellular immune receptors that recognize, directly or indirectly, specific pathogen effectors encoded by avirulence (*Avr*) genes (Bent and Mackey 2007). For example, in solanaceous plants, four *R*-genes were found to detect *Avr* proteins directly, while four others recognized the *Avr* protein indirectly, which exerted effects on a target protein (van Ooijen et al. 2007).

The deduced NBS-LRR proteins can be divided into two subfamilies, TIR-NBS-LRR (TNL) and non-TNL based on their N-terminal features. TNLs possess a domain with similarity to both the intracellular signaling domains of *Drosophila* Toll and the mammalian Interleukin-1 receptor (TIR). Non-TNL often have a putative N-terminal coil-coil (CC) and are thus designated CNLs (Dangl and Jones 2001). The highly conserved NBS domain in the *R* proteins

has been demonstrated to be able to bind and hydrolyze ATP and GTP and this domain is also found in other prokaryotic and eukaryotic proteins (Meyers et al. 1999). Different from NBS domain, the LRR motif is typically involved in protein-protein interactions and responsible for recognition specificity (Dangl and Jones 2001).

Gene encoding NBS-LRR proteins comprise one of the largest gene families in plant genomes, with a total of 149 and 480 members identified in *Arabidopsis thaliana* and *Oryza sativa*, respectively (Meyers et al. 2003; Zhou et al. 2004; Yang et al. 2006). In addition, CNLs are found widely in angiosperm (Cannon et al. 2002). In rice, three atypical "TNs" (TNL but lacking an LRR domain) have been characterized, which differ substantially from typical TNL genes (Bai et al. 2002). These data suggested that *R*-gene arsenals have diverged substantially between monocot and dicot lineages over the course of evolution. More recently, approximately 330 NBS-encoding genes were detected in incomplete genome sequences (~58% of the whole genome) of *Medicago truncatula* (Ameline-Torregrosa et al. 2008). In addition, over 1,600 NBS sequences have been amplified from a diverse array of plant species using degenerate PCR primers (Cannon et al. 2002; McHale et al. 2006). Furthermore, results of nucleotide polymorphism analyses have demonstrated extremely high levels of inter- and intraspecific variation, which presumably evolved rapidly in response to pathogen population changes (Ding et al. 2007a, b; Shen et al. 2006; Yang et al. 2007).

Most NBS-LRR genes are unevenly distributed in plant genomes and existing mainly as multi-gene clusters (Meyers et al. 2003; Zhou et al. 2004; Ameline-Torregrosa et al. 2008). In the rice genome, 24.98% *R*-like genes are present on chromosome 11, with most located in several large clusters. These genes were possibly originated by tandem duplication and subsequent divergence under the selective pressure of rice pathogens (Rice Chromosomes 11 and 12 Sequencing Consortia 2005). The clustered distribution of *R*-genes provides a reservoir of genetic variation from which new specificities to pathogens can evolve via gene duplication, unequal crossing-over, ectopic recombination or diversifying selection (Michelmore and Meyers 1998). However, the specific evolutionary routes of NBS-encoding genes still remain elusive. The complete genome sequences of grapevine (*Vitis vinifera*), poplar (*Populus trichocarpa*), *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*) provide an opportunity to conduct comparative genomic analyses of NBS-encoding genes (*Arabidopsis* Genome Initiative 2000; International Rice Genome Sequencing Project 2005; Jaillon et al. 2007; Tuskan et al. 2006). In this study, we first identified 535 and 416 respective NBS-encoding genes residing in grapevine and poplar genomes. We subsequently analyzed their evolution and genomic organization. Different patterns of TIR versus non-TIR-NBS-encoding gene evolution were observed. This analysis also suggested that recent tandem duplication could play a major role in the expansion of NBS-encoding genes in the two woody species.

Materials and methods

Identification of NBS-encoding genes

Grapevine (*Vitis vinifera*) assembly and annotation V1.0 were downloaded from http://www.genoscope.cns.fr/externe/English/Projets/Projet_ML/index.html. Poplar (*Populus trichocarpa*) genome sequences V1.1 were found in http://www.jgi.doe.gov/poplar. The same methods (as in Zhou et al. 2004; Yang et al. 2006) were employed to identify NBS-encoding genes in grapevine and poplar. Firstly, amino acid sequence of NB-ARC domain (Pfam: PF00931) was adopted as a query in BLASTP searches for possible homologs encoded in grapevine and poplar genomes. The threshold expectation value was set to 10^{-4}, a value determined empirically to filter out most of the spurious hits. Secondly, the NBS domain was determined by Pfam version 22.0 (http://pfam.janelia.org/). Then, the nucleotide sequences of candidate NBS genes were used as queries to find homologs in grapevine or poplar genomes by BLASTn searches. This step was crucial to find the maximum number of candidate genes. All 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 (http://www.softberry.com/) and GENSCAN (http://genes.mit.edu/genescan.html/) to obtain information on complete open reading frames (ORFs).

To exclude potentially redundant candidate NBS genes, all sequences were orientated by BLASTn, and sequences located in the same location were eliminated. All non-redundant candidate NBS genes were surveyed to further verify whether they encoded TIR, CC, NBS, or LRR motifs using the Pfam database (http://pfam.janelia.org/), SMART protein motif analyses (http://smart.embl-heidelberg.de/), MEME (Multiple Expectation Maximization for Motif Elicitation, Bailey and Elkan 2005), and COILS with a threshold of 0.9 to specifically detect CC domains (Lupas et al. 1991). This detailed information on protein motifs and domains was used to classify the NBS-encoding genes into subgroups (Table 1).

Sequence alignment and phylogenetic analysis

Multiple alignments of amino acid sequences were performed using ClustalW with default options (Thompson et al. 1994). The alignments were carried out using the

Table 1 The number of genes that encode domains similar to NBS genes in four plant genomes

Predicted protein domains	Letter code	Grapevine	Poplar	*Arabidopsis*[a]	Rice[b]
Total NBS-encoding genes		535	416	174	519
NBS-LRR type genes					
TIR-NBS-LRR	TNL				
TIR-NBS-LRR	TNL′	90	73	93	0
TIR-CC-NBS-LRR	TCNL	1	0	0	0
TIR-NBS-LRR-TIR	TNLT	3	4	0	0
TIR-NBS-LRR-TIR-NBS	TNLTN	1	1	0	0
TIR-NBS-LRR-TIR-NBS-LRR	TNLTNL	2	0	0	0
Non-TIR-NBS-LRR	Non-TNL				
CC-NBS-LRR	CNL	200	119	51	159
CC-NBS-NBS-LRR	CNNL	2	1	0	1
CC-NBS-LRR-NBS-LRR	CNLNL	1	0	0	0
X-NBS-LRR	XNL	147	132	0	304
NBS_{CC}-LRR	NL	12	0	3	0
Total		459	330	147	464
NBS type genes					
CC-NBS	CN	26	14	5	7
X-NBS	XN	36	62	1	45
TIR-NBS	TN	14	10	21	3
Total		76	86	27	55
TIR-X type genes	TX	27	108	30	1

CC coiled-coil domain, *LRR* leucine-rich repeat domain, *NBS* nucleotide-binding site, *TIR* Toll/interleukin-1-receptor, *X* some unknown motifs

[a] Data from Meyers et al. (2003)

[b] Data from Zhou et al. (2004) and Yang et al. (2006)

NBS region, based on the Pfam results. Phylogenetic trees were constructed based on the Bootstrap neighbor-joining (NJ) method with a Kimura 2-parameter model by MEGA v4.0 (Tamura et al. 2007). The stability of internal nodes was assessed by bootstrap analysis with 1,000 replicates.

The number of nonsynonymous substitutions per nonsynonymous site and the number of synonymous substitution per synonymous site were denoted by K_a and *Ks* respectively. In order to detect the mode of selection, we evaluated the ratio of nonsynonymous to synonymous nucleotide substitutions (K_a/K_s) among paralogs. Protein sequences of paralogs were initially aligned by ClustalW, and the resulting alignments were then used to guide the aligning of nucleotide coding sequences (CDSs) using MEGA. The K_a and K_s were calculated by DnaSP v4.0 (Rozas et al. 2003) based on Nei and Gojobori (1986). Generally, a K_a/K_s ratio >1 indicates positive selection, the ratio <1 indicating negative or purifying selection, while the ratio 1 indicating neutral evolution. Full-length CDS sequences were used to infer the information of gene duplication. The time of duplication event for each node in phylogenetic trees was determined by average K_s of the node, which was calculated by taking the arithmetic mean of all left-right branch combinations (Fig. 1). For example, K_s for node AII was determined by the sum of the K_s of sequence pairs A&C and B&C divided by the number of pairs; K_s for node BIII was determined by the sum of the K_s of sequence pairs A&C, A&D, B&C and B&D divided by the number of pairs (Fig. 1).

Nucleotide divergence among paralogs was estimated by π with the Jukes and Cantor correction (Lynch and Crease 1990) using DnaSP v4.0 (Rozas et al. 2003). A gene cluster was defined as a region in which two neighboring homologous genes were < 200 kb apart (Holub 2001). Sequence exchange between duplicates was investigated by the program GENECONV 1.81 (http://www.math.wustl.edu/~sawyer/geneconv/). The default setting with 10,000 permutations was used for the analysis. The statistical significance of gene conversion events were defined as a global permutation *P* value of <0.05.

Results

Total number of NBS-encoding genes in grapevine and poplar

Five hundred thirty-five and 416 NBS-encoding genes, including 459 and 330 NBS-LRR-encoding genes, were identified in grapevine and poplar, respectively (Table 1). Among the 459 NBS-LRR grapevine genes, 97 were identified as TNL resistance genes, including TNL′, TCNL, TNLT, TNLTN, and TNLTNL (Table 1) and 362 non-TNLs, including CNL, CNNL, CNLNL, XNL, and NL. In addition, 76 NBS type genes, which lacked LRR domains, were detected containing CN, TN and XN genes (Table 1). In the poplar genome, 78 and 252 respective NBS-encoding

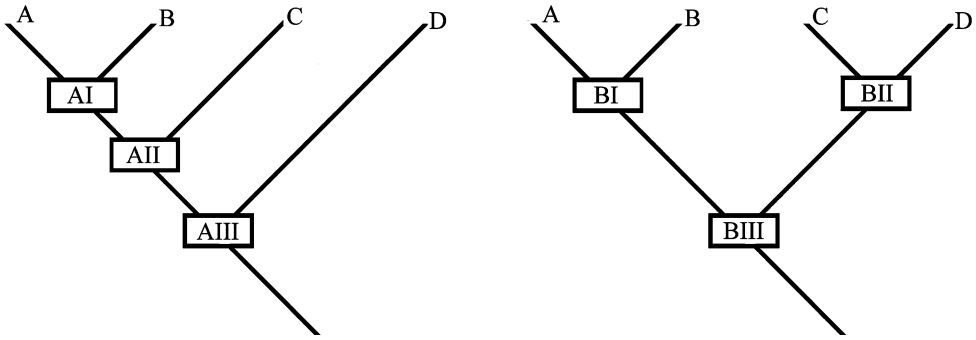

Fig. 1 Schematic diagram average K_s calculations at each node, each node represents one duplication event. The K_s value for each node was calculated by taking the arithmetic mean of all left-right branch combinations. For example, K_s for node AII was determined by the sum of K_s sequence pairs A&C and B&C divided by the number of pairs; K_s for node AIII was determined by the sum of K_s sequence pairs A&D, B&D and C&D divided by the number of pairs; K_s for node BIII was determined by the sum of K_s sequence pairs A&C, A&D, B&C and B&D divided by the number of pairs

genes were classified as TNLs and non-TNLs. Eighty-six NBS genes were found, which lacked the LRR domain (Table 1).

Approximately 30,434, 45,555, 27,000 and 37,544 protein-coding genes were estimated in the full sequenced grapevine, poplar, *Arabidopsis* and rice genomes, respectively (Jaillon et al. 2007; Tuskan et al. 2006; *Arabidopsis* Genome Initiative 2000; International Rice Genome Sequencing Project 2005). The NBS-LRR genes accounted for approximately 1.51, 0.72, 0.53 and 1.23% of all predicted ORFs in these four species, respectively. Both the absolute number and relative proportion of NBS-LRR genes in grapevine and rice were significantly higher than those in poplar and *Arabidopsis* genomes (Table 1).

The average number of exons detected in NBS-LRR genes was 3.96 in grapevine, 3.72 in rice, 2.35 in poplar and 4.19 in *Arabidopsis.* These exon estimates were less than that in all predicted genes within each of the four taxa, which was 5.0, 4.9, 4.3 and 5.8 per gene, respectively (Jaillon et al. 2007; International Rice Genome Sequencing Project 2005; Tuskan et al. 2006; *Arabidopsis* Genome Initiative 2000). The average TNL exon numbers were significantly greater than those in non-TNLs (*t* test, $P < 0.001$). In grapevine, 7.68 TNL exons were detected *vs.* 3.22 non-TNLs and 3.5 vs. 2.23 in poplar. This result was consistent with previous studies in *Arabidopsis* that indicated most CNL genes were encoded by a single exon (2.17 on average) and TNLs were encoded by multiple exons (5.25 on average; Meyers et al. 2003).

Organization and phylogeny of NBS-encoding genes in two woody genomes

Studies in *Arabidopsis* and rice report an uneven distribution of NBS-encoding genes on chromosomes (Meyers et al. 2003; Zhou et al. 2004). In the present study, based on the location of individual NBS genes on chromosomes (Jaillon et al. 2007; Tuskan et al. 2006), 354 out of 535 grapevine NBS-encoding genes (66.2%) were mapped on 19 chromosomes. The remaining genes were situated on unanchored supercontigs (chromosome(s) unknown). For the 354 anchored NBS genes, 236 (66.7%) were physically located on chromosomes 9, 12, 13 and 18, with 49, 49, 58 and 80 NBS-encoding genes on each respective chromosome. NBS genes were absent from chromosomes 7 and 10 (Table 2; Fig. S1). Similarly, in poplar 43.3% (180 out of 416) of NBS genes were mapped on 19 chromosomes (Table 2; Fig. S2); 61.7% (111 out of 180) were found physically localized on chromosomes 1, 2, 3 and 19, including 26, 23, 20 and 42 genes, respectively (Table 2; Fig. S2). In addition, two chromosomes lacked NBS-encoding genes, e.g. chromosomes 10 and 17.

The majority of NBS-containing genes in the grapevine (83.2%) and the poplar (67.5%) genomes were found in clusters (Tables 2 and 3; Fig. S1 and S2). A gene cluster is defined as a region in which two neighboring homologous genes are <200 kb apart. In grapevine, 77 clusters, including 445 NBS genes, were identified (Tables 2 and 3). On an average, 5.78 NBS members were detected in a cluster, which could be slightly underestimated due to unanchored scaffold sequences. The largest cluster, comprised of 26 NBS members (14 TNLs, 7 TNs and 5 XNLs), was detected on chromosome 18 (Fig. S1). In poplar, there were 281 NBS genes in 75 clusters (Table 2), with an average of 3.75 NBS members per cluster. The largest cluster, containing 19 NBS members (4 TNLs, 2 CNLs, 9 XNLs and 4 XNs), was found on chromosome 19.

Previous studies have shown that phylogenies constructed from the NBS domain were robust and distinguished TNL and non-TNL subfamilies (Meyers et al. 1999; Meyers et al. 2002). These two distinct groups (Bootstrap value >98%) were also identified in the phylogenetic reconstruction using grapevine and poplar conserved NBS protein sequences (Figs. S3 and S4). In grapevine, the

Table 2 Distribution of NBS-encoding *R*-genes on chromosomes in grapevine and poplar genomes

Chromosome	Species	TNL	TN	CNL	CN	XNL	XN	Clusters (Gene no.)	Total	Tir_X
1	Grapevine	2	0	6	0	1	1	2 (10)	10	0
	Poplar	4	1	10	3	3	5	6 (18)	26	2
2	Grapevine	0	0	0	0	1	0	0 (0)	1	0
	Poplar	1	0	16	0	4	2	3 (20)	23	1
3	Grapevine	0	0	5	1	6	2	1 (10)	14	0
	Poplar	2	1	9	0	5	3	3 (12)	20	3
4	Grapevine	0	0	1	0	0	0	0 (0)	1	0
	Poplar	1	0	0	0	1	0	1 (2)	2	1
5	Grapevine	11	1	2	0	11	0	4 (25)	25	1
	Poplar	1	0	1	0	1	1	1 (2)	4	1
6	Grapevine	2	0	4	0	6	1	4 (12)	13	1
	Poplar	0	0	3	0	2	0	1 (3)	5	0
7	Grapevine	0	0	0	0	0	0	0 (0)	0	0
	Poplar	3	0	3	0	1	0	2 (5)	7	1
8	Grapevine	0	0	1	0	1	0	1 (2)	2	0
	Poplar	2	0	0	0	0	0	1 (2)	2	0
9	Grapevine	0	0	32	3	7	7	6 (45)	49	0
	Poplar	0	0	0	0	1	1	0 (0)	2	0
10	Grapevine	0	0	0	0	0	0	0 (0)	0	0
	Poplar	0	0	0	0	0	0	0 (0)	0	0
11	Grapevine	0	0	4	0	2	0	1 (3)	6	1
	Poplar	5	3	4	1	3	1	2 (16)	17	14
12	Grapevine	13	1	13	1	13	7	6 (44)	49	0
	Poplar	2	0	3	1	3	0	1 (5)	9	0
13	Grapevine	1	0	27	3	23	4	7 (54)	58	0
	Poplar	4	0	4	0	3	2	2 (5)	13	1
14	Grapevine	0	0	5	0	1	1	2 (5)	7	0
	Poplar	0	0	1	0	0	0	0 (0)	1	0
15	Grapevine	0	0	6	2	4	1	3 (10)	13	0
	Poplar	0	1	1	0	1	0	1 (2)	3	0
16	Grapevine	0	0	0	0	2	0	0 (0)	2	0
	Poplar	0	0	0	0	1	0	0 (0)	1	1
17	Grapevine	0	0	3	0	2	0	1 (3)	5	0
	Poplar	0	0	0	0	0	0	0 (0)	0	0
18	Grapevine	54	8	4	0	13	1	8 (76)	80	11
	Poplar	1	0	1	0	1	0	1 (2)	3	1
19	Grapevine	0	0	11	1	6	1	4 (16)	19	2
	Poplar	11	0	2	0	22	7	7 (41)	42	11
Chro-Unknown	Grapevine	14	4	79	14	60	10	27 (130)	181	11
	Poplar	41	4	62	9	80	40	43 (146)	236	71
Total	Grapevine	97	14	203	26	159	36	77 (445)	535	27
	Poplar	78	10	120	14	132	62	75 (281)	416	108

TIR-NBS clade included 97 TNLs, 14 TNs, 2 CNLs, 13 XNLs and 1 XN. Interestingly, 2 CNL genes, *G_11213* (on unknown chromosome) and *G_21959* (on chromosome 18), were found nested in the TIR-NBS clade. *G_11213* showed a high nucleotide identity (86%) with a typical TCNL gene (*G_38789*). Compared to *G38789*, a 90 amino acid deletion was found at the N-terminal end of *G_11213*, suggesting a truncation or fusion of domains leading to novel domain compositions. The other CNL gene resided in a cluster containing 13 TNLs or TNs members, suggesting TN ancestry.

Table 3 Organization of NBS-encoding *R*-genes in the four plant genomes

Organization	Grapevine	Poplar	*Arabidopsis*	Rice
Single-genes	119 (172)	91 (106)	93 (134)	240 (460)
Multi-genes	416 (363)	325 (310)	81 (40)	279 (59)
Gene family no.	94 (102)	61 (64)	25 (15)	93 (24)
Maximal members of a family	13 (10)	23 (17)	7 (4)	10 (3)
Average members per family	4.43 (3.56)	5.33 (4.84)	3.24 (2.67)	3.00 (2.46)
Singleton genes	90	135	46	157
Clustered genes	445	281	125	362
Cluster no.	77	75	39	104
Maximal members of a cluster	26	19	11	11
Average members per cluster	5.78	3.75	3.21	3.48

A multi-gene family was defined as genes within >70% (or >80%) amino acid identify

In addition, numerous XNLs or XNs were located in the TIR-NBS clade, indicating that all the XNL or XN genes might be TIR-NBS-related, although no TIR domain was detected.

In grapevine, 80.4% of TNLs and 71.4% of TNs were located on chromosomes 5, 12 and 18 (Table 2; Fig. S1). Chromosome 18 was nearly an exclusive TIR-NBS chromosome (Table 2), with 55.7% (54 out of 97) of the TNLs and 57.1% (8 out of 14) of the TNs were detected on this chromosome. Furthermore, 14 NBS genes, 13 XNLs and 1 XN, were found on this chromosome. These genes were positioned in the TIR-NBS clade and therefore were most likely TIR-related (Fig. S3). The remaining four genes on chromosome 18 were CNLs (Table 2). The phylogenetic analysis suggested that *G_11213* (as mentioned above) evolved from a TNL gene with a TIR domain deletion. Two of the other three CNLs, *G_01807* and *G_21959*, were single genes located in heterogeneous clusters mixing with TIR-NBS genes. The remaining 1 CNL had a high nucleotide identity with a large NBS gene family (identity ranging from 77 to 95% and 87% on average) that included other 8 CNL and XNL members located on chromosome 5, suggesting that this gene might translocate to chromosome 18 from chromosome 5 after gene duplication.

The poplar phylogeny (Fig. S4) shows the TIR-NBS clade were composed of 78 TNLs, 10 TNs, 18 XNLs and 15 XNs. The non-TIR-NBS clade exhibited an intriguing phylogenetic subclade (bootstrap value >95%), which included 59 members of 45 XNLs, 3 CNLs and 11 XNs. In this subclade, 45 members carried an N-terminal BED binding zinc finger domain (Zf_BED domain; Pfam: PF02892; Tuskan et al. 2006), containing 35 Zf_BED-NBS-LRRs (BNL), 2 Zf_BED-CC-NBS-LRRs (BCNL) and 8 Zf_BED-NBS lacking LRR (BN). However, none of the Zf_BED-NBS genes could be detected outside the poplar clade, with the exception one BNL (*Xa1*) in rice.

Another domain (Pfam: PF05659) shared by the broad-spectrum resistance protein *RPW8* and its family members in Brassicaceae (Xiao et al. 2001, 2004) was detected in a few NBS genes from grapevine and poplar (Figs. S3 and S4). Five CNLs or CNs (*G_08450*, *G_03901*, *G_10062*, *G_04842*, and *G_12615*) in grapevine and one XNL (*P_563015*) in poplar were found carrying this domain. This result suggested that NBS-LRR novel domain compositions might occur by the fusion of CNL and RPW8 domains. Therefore, the *RPW8* gene exhibiting broad-spectrum mildew resistance in *Arabidopsis* could have ancient origins and have co-evolved with or possess functional requirements to NBS-LRR genes.

NBS gene duplication

Based on the NBS-LRR gene analysis in *Arabidopsis*, results suggested that both tandem and large-scale block duplication contributed to the expansion of this gene group (Leister 2004). To identify duplicated gene pairs, we defined a gene family according to the following criteria (Zhou et al. 2004): (1) the alignable nucleotide sequence covered >70% of the longer aligned gene, and (2) the amino acid identity between the sequences was >70%. Ninety-four and 61 NBS gene families were identified in grapevine and poplar, including 416 (77.8%) and 325 (78.1%) respective NBS genes. The percentage of NBS genes in multi-gene families (two or more members in a family) of the two woody species was significantly higher than those (47.3 and 53.7%) in the *Arabidopsis* and rice genomes (Table 3). The average number of NBS members per multi-gene family was 4.43 (ranging from 2 to 10) and 5.33 (ranging from 2 to 23) in grapevine and poplar, respectively. Furthermore, the size of the members was significantly larger than those (3.24 and 3.00) in *Arabidopsis* and rice (Table 3). This analysis revealed more duplication in the genomes and the multi-gene families of grapevine and poplar.

To distinguish recent duplications of NBS genes in plant genomes, we employed a more stringent criteria for defining a multi-gene family, e.g. an amino acid identity >80%. Under this rule, 23.4 and 11.4% of the NBS genes belonged to multi-gene families in *Arabidopsis* and rice, respectively

(Table 3). However, 67.9 and 74.5% of the respective multi-genes were identified in grapevine and poplar, comparable to the percentages generated with a 70% amino acid identity criterion. These results suggested recent duplications were common in the two perennial species and the NBS phylogenetic trees confirmed the recent duplications. Short branches from the terminal nodes were more frequent in grapevine and poplar than those in *Arabidopsis* and rice (Fig. S5).

The time of duplication events was estimated by the average K_s at each node in the phylogenetic tree. We constructed phylogenetic trees using full-length protein sequences from each genome, and divided them down into many smaller phylogenetic units using a >70% amino acid identity. In each phylogenetic unit, every node represented one duplication event (Fig. 1). The average K_s of each node was calculated by taking the arithmetic mean of all left-right branch combinations (Fig. 1). K_s was considered as the time proxy for duplication event, and the frequency distributions of individual K_s values reflected the relative time of genome duplications. A quite different individual K_s distribution was revealed between perennial and annual species (Fig. 2). Compared with both *Arabidopsis* and rice genomes, a higher proportion of young duplicates (the smallest K_s classes) were present in grapevine and poplar, indicating recent expansion. Additional analyses indicated that 91.8% of NBS multi-genes and all duplicates with $K_s <$ 0.1 were located in tandem clusters in grapevine. In poplar, the proportion of those genes within tandem clusters was 72.9 or 84.6%, respectively. The clustered locations of young duplicates provided additional evidence that recent tandem duplication, but not ancient large-scale duplication played the major role in NBS gene expansion in the two perennial species.

TIR and non-TIR NBS-encoding genes evolution

Phylogeny reconstruction was employed to estimate evolutionary patterns of TIR- and non-TIR-NBS-encoding genes using only conserved NBS domains of the perennial and annual species (Fig. S5). Both TIR- and non-TIR-group were identified in all genes (Fig. S5), therefore we subsequently constructed separate phylogenies for the two subgroups (Figs. 3 and 4).

The analysis revealed different evolutionary patterns of TIR- and non-TIR-NBS-encoding genes. The TIR-NBS

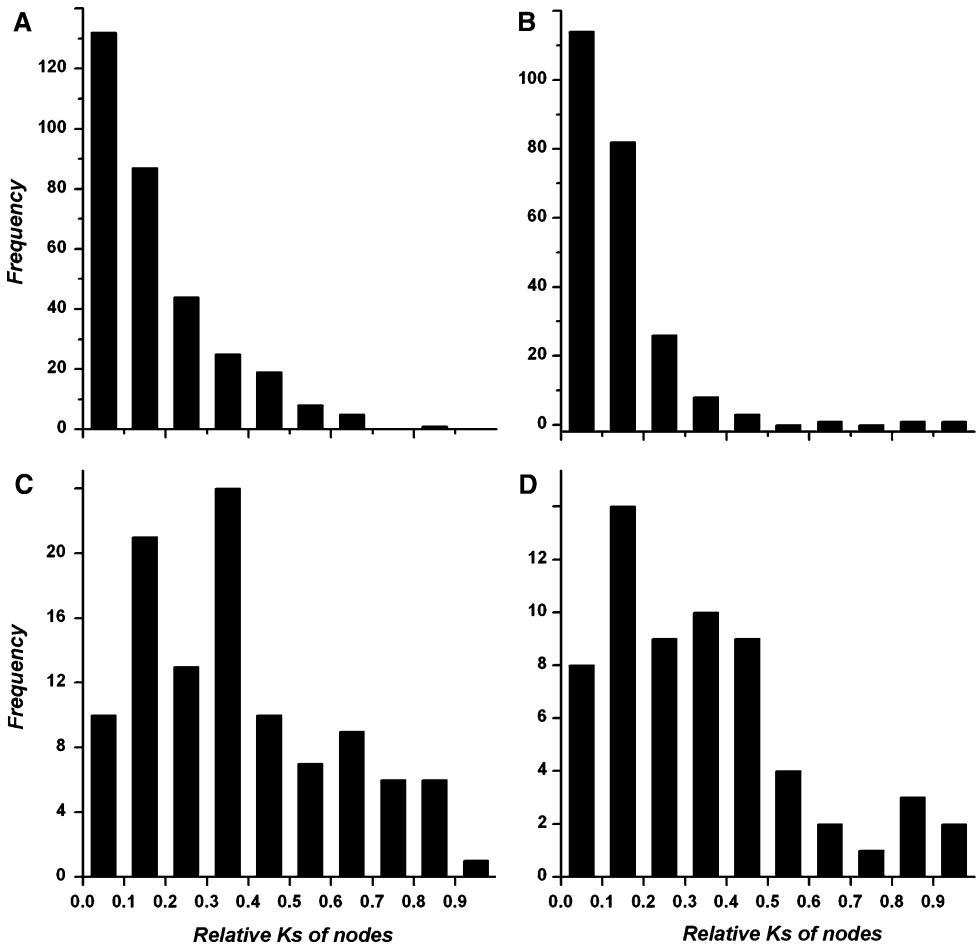

Fig. 2 The frequency distribution of relative K_s nodes in the four plant genomes. **a** grapevine; **b** poplar; **c** rice; **d** *Arabidopsis*

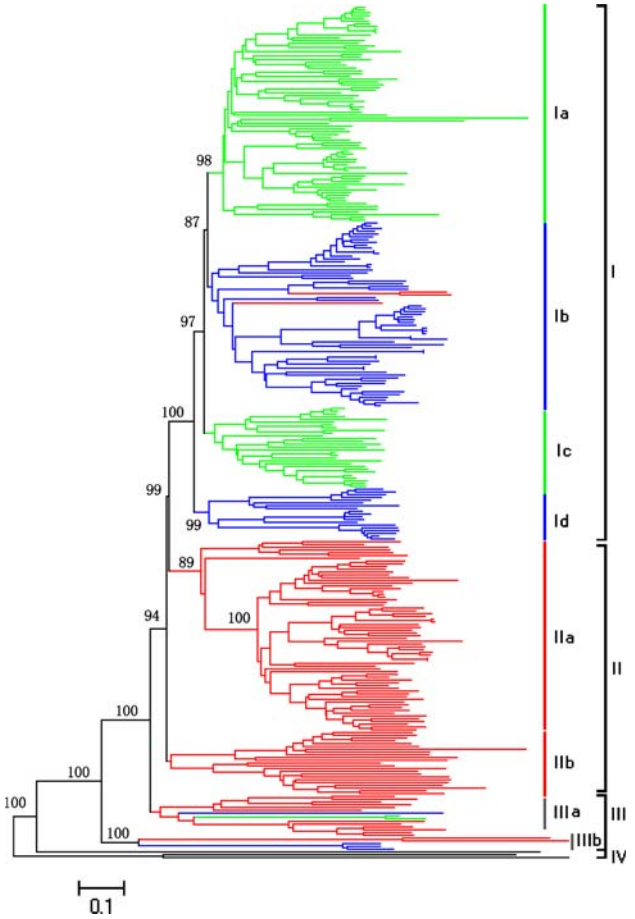

Fig. 3 Phylogenetic tree based on NBS domains from TIR-type NBS genes. *Green* (grapevine), *Blue* (poplar), *Red* (*Arabidopsis*), and *Black* (rice). Subclade I represents a woody-specific clade; subclade II represents an *Arabidopsis*-specific clade; subclade III represents a mixed clade; and subclade IV represents a rice-specific clade

phylogenetic tree was divided into four major clades supported by branch lengths and bootstrap values (>80%; Fig. 3). Clade I was largely a woody-specific clade, with grapevine-specific subclades Ia and Ic and poplar-specific subclades Ib and Id. However, subclade Ib also contained three TNLs from *Arabidopsis*. Clade II was an *Arabidopsis*-specific clade including two distinct subclades. Clade III was also comprised of two distinct subclades but was represented by genes from all three dicot species. Clade IV was represented by three atypical TNs from rice (Bai et al. 2002) and was therefore considered as the outgroup. The phylogenetic distribution in this tree exhibited a lineage-specific expansion in the TIR-type NBS genes, suggesting these genes might be traced to a small number of species-specific progenitors.

In contrast, the non-TIR-NBS clades contained NBS genes from at least two or more species (Fig. 4). Nevertheless, three distinct species-specific clades could be detected. Subclade A and B were rice-specific with 106 and 181 NBS genes, and subclade C was a Zf_BED-NBS poplar specific

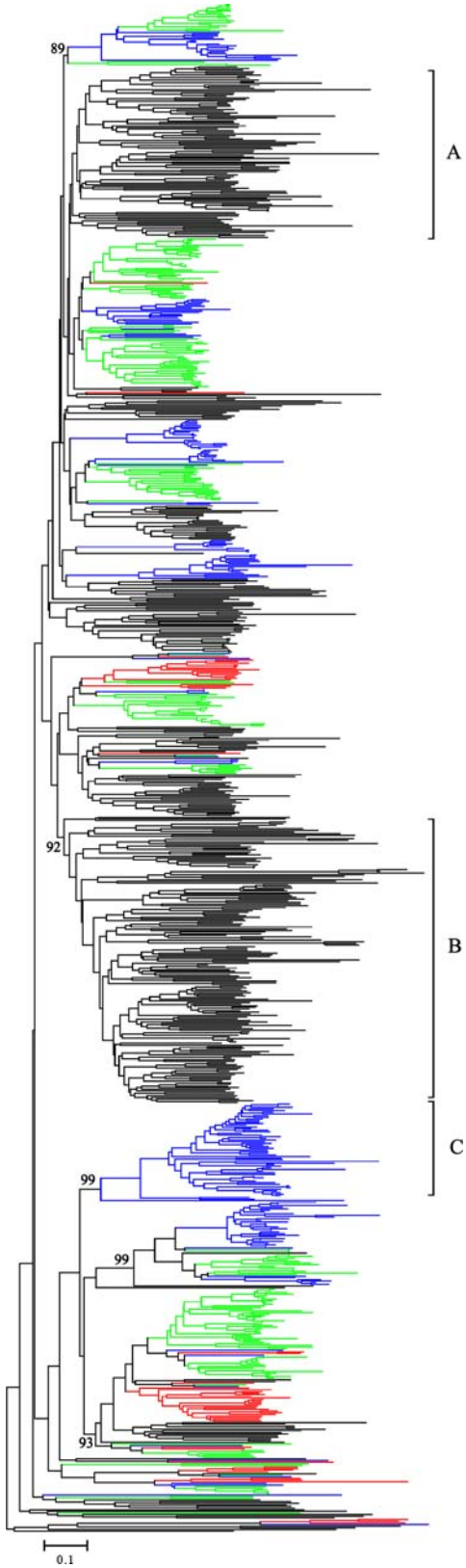

Fig. 4 Phylogenetic tree based on NBS domains from non-TIR-type NBS genes. *Green* (grapevine), *Blue* (poplar), Red (*Arabidopsis*), and *Black* (rice). Subclades A and B represent rice-specific clades; and subclade C represents a poplar-specific BED-NBS clade

clade comprised of 59 members. In general, bootstrap values were low for the non-TIR-NBS analysis, resulting in equivocal results. Results revealed highly divergent non-TIR-NBS genes suggesting a large number of non-specific progenitors.

In summary, the topologies of the TIR and non-TIR trees were significantly different. Species-specific sequences resulted in short branch lengths in the TIR subfamily. However, multiple sequences resulted in several divergent cladeds in the non-TIR-NBS subfamily. Furthermore, orthologous genes were readily identified in some non-TIR-NBS clades, while orthologs were not detected in the TIR-NBS phylogeny.

Duplicated NBS gene non-synonymous to synonymous substitution and gene conversion

To explore different selective constrains on duplicated NBS genes, the K_s and K_a/K_s ratio for each pairwise of duplicates were calculated. We excluded the duplicated pairs with a $K_s < 0.005$, which was generally too small to obtain a reliable K_a/K_s ratio (Zhang et al. 2003). K_s was used as a proxy for time and resulted in a significant negative correlation between K_a/K_s and K_s among the most closely related paralogs in both perennial and annual taxa (Fig. 5a). These results showed a significantly higher K_a/K_s in the recent NBS duplicates, and suggested positive selection or relaxation of negative selection occurred soon after duplication events. The RLK gene family demonstrated similar results consistent with the expectation that higher levels of nonsynonymous polymorphisms occurred in recently duplicated genes as a result of functional redundancy (Shiu et al. 2004).

The K_a/K_s ratio varied in different individual NBS-LRR gene domains (Jiang et al. 2007). Our duplicated gene family analysis in the two woody species confirmed that the ratio of the LRR domains was significantly higher than the NBS regions (Fig. 5 b, c; a two-tailed t test, $P < 0.001$). The average K_a/K_s ratios for the LRR and the NBS regions were 0.666 (ranging from 0.22 to 2.78) and 0.552 (0.27–1.46) in grapevine, and 0.733 (0.21–1.49) and 0.518 (0.08–0.96) in poplar, respectively. Relative to the NBS region, a higher K_a/K_s ratio in the LRR region was detected in grapevine (72.7% of the families) and poplar (68.1% of the families). In addition, the number of the LRR regions demonstrating a $K_a/K_s > 1$ exceeded that of NBS regions.

Interestingly, the K_a/K_s ratio was significantly higher in the non-TIR NBS families than the TIR NBS families (0.664 vs. 0.585 in poplar and 0.673 vs. 0.575 in grapevine; $P < 0.05$). This observation was consistent with previous studies (Cannon et al. 2002), suggesting stronger diversifying selection in the non-TIR-NBS subfamily.

Gene conversion (and/or unequal crossing-over) events were identified using the GENECONV program. In total, 823 and 468 gene conversion events, involving 299 and 187 NBS-encoding genes, were detected in grapevine and poplar, respectively. However, only 143 and 81 gene

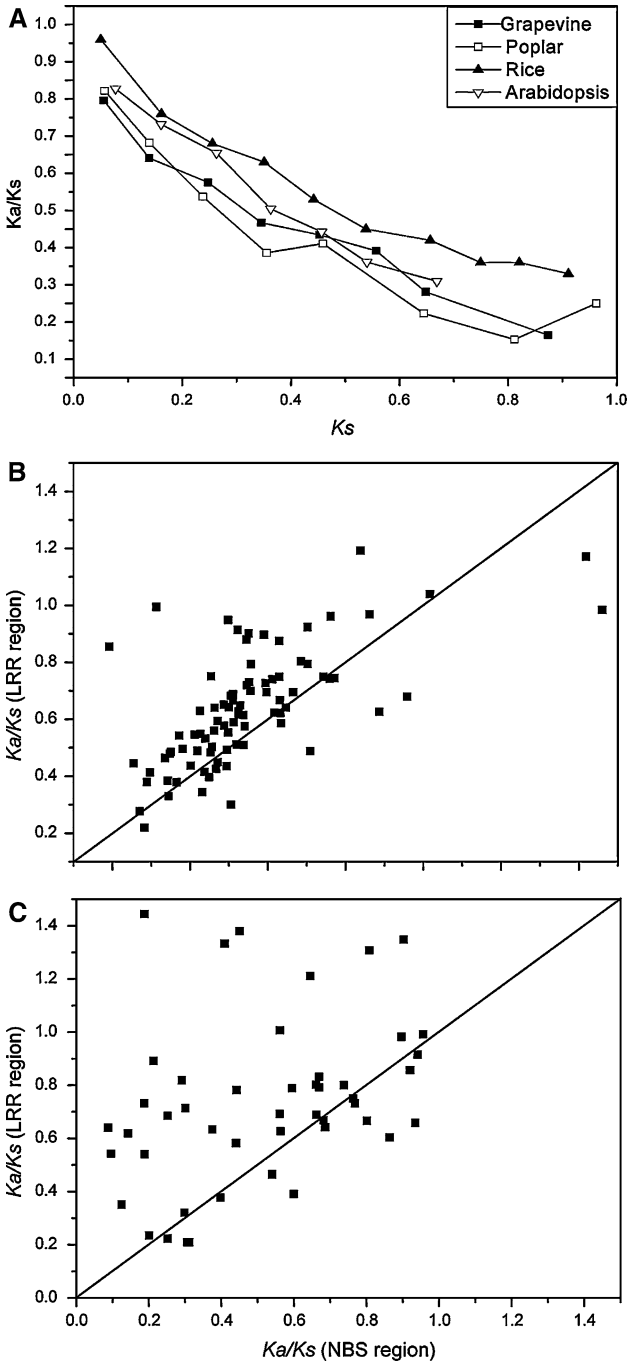

Fig. 5 **a** The relationships between K_a/K_s and K_s NBS paralogs in grapevine ($r^2 = 0.97$, $P < 10^{-4}$), poplar ($r^2 = 0.84$, $P < 0.005$), *Arabidopsis* ($r^2 = 0.93$, $P < 10^{-4}$) and rice ($r^2 = 0.98$, $P < 10^{-4}$), respectively. The *X*-axis denotes average K_s per unit of 0.1, e.g. K_s = 0–0.1, 0.1–0.2, 0.2–0.3 etc. The *Y*-axis denotes average K_a/K_s ratios. (**b**) and (**c**) K_a/K_s ratio in different domains in grapevine (**b**) and poplar (**c**). Each *dot* denotes the average K_a/K_s ratio in a NBS family. The *X*-axis and Y-axis denote average K_a/K_s in the NBS and LRR region, respectively. The *line* indicates equal K_a/K_s ratios between NBS and LRR regions

conversion events were detected in *Arabidopsis* and rice, indicating more sequence exchange between NBS-encoding paralogs in perennial species than those in annual plants.

Discussion

Recent tandem duplication in NBS gene expansion of perennial species

Previous studies identified 174 and 519 NBS-encoding genes, including 149 and 464 NBS-LRR-encoding genes in the *Arabidopsis* and rice genomes, respectively (Meyers et al. 2003; Zhou et al. 2004; Yang et al. 2006). Gene duplication, including tandem and whole-genome duplication (WGD), has resulted in a substantial increase in gene number. However, NBS-LRR genes in WGD blocks were <10% in *Arabidopsis* and 0% in rice (Nobuta et al. 2005; Yu et al. 2005). In poplar, WGD resulted in two duplication blocks containing 12 NBS genes. These results were determined by shared phylogenetic clade combinations within multiple physical clusters and by localizations within regions of demonstrated intragenomic synteny (Figs. S2 and S4). WGD NBS blocks were not detected in grapevine (Figs. S1 and S3), indicating that soon after WGD, most duplicated NBS genes were lost. On the other hand, ~91.8, 72.9, 71.2 and 74.6% of NBS family members were detected in tandem clusters of grapevine, poplar, *Arabidopsis* and rice, respectively. These results suggested that tandem duplication, but not WGD, played a major role in NBS gene expansion in the four plant genomes. Tandemly clustered *R*-genes distribution could provide a reservoir of genetic variation from which new disease resistant specificities could evolve (Michelmore and Meyers 1998).

However, the duplications shapes in the two woody species were significantly different from those in rice and *Arabidopsis* based on K_s as a time proxy for duplication events (Fig. 2). In grapevine and poplar, the recent tandem duplications (K_s = 0–0.1) played a substantial role in NBS gene expansion (42.3% in grapevine and 49.4% in poplar). However, a majority of tandem duplicates were found for the K_s 0.1–0.4 time frames in *Arabidopsis* (54.6%) and rice (55.0%). Sequence exchanges between the NBS paralogs also showed that increased gene conversion events (about 3–10-fold larger) were detected in the two perennial species compared with the annual taxa. These results suggested divergent evolutionary patterns between the annual and perennial plants investigated in this study.

The likelihood that a perennial plant will encounter a pathogen or herbivore before reproduction is unequivocal (Tuskan et al. 2006). Compared with the annual plants included in this study, e.g. rice and *Arabidopsis*, the life history of long-lived perennials limits the plants to catch the evolutionary rates of a microbial or insect pest (Tuskan et al. 2006). Therefore, the recent frequent tandem duplications in woody perennials could result in an increased probability for meiotic and somatic recombination to evolve more novel disease resistance specificities (Parniske et al. 1997).

The K_a/K_s ratio declined sharply over time when using K_s as a time proxy in the four plant genomes (Fig. 5a), suggesting positive selection or relaxation of negative selection on the newly duplicated *R*-genes. The results also indicated that the majority of duplicated NBS-encoding genes might accumulate degenerative mutations (under a relaxed purifying selection or positive selection) for a period of time, followed by functional specialization (under purifying selection again) by complementary partitioning of ancestral or newly evolved functions.

Unique evolutionary patterns between TIR- and non-TIR-NBS genes

The conserved NBS domain was used to study the genomic architecture of this gene family. The TIR and non-TIR sequences, including the three atypical rice TNs, were deeply separated in the four plant species (Fig. S5), suggesting ancient origins and subsequent divergence between the two NBS genes types. The number of TNLs was similar in *Arabidopsis*, poplar and grapevine (93, 78 and 97, respectively; Table 1). However, the number of non-TIR-NBS-LRR genes varied from 54 in *Arabidopsis*, 252 in poplar, 362 in grapevine and 464 in rice. Therefore, within each of the four genomes, the total number of *R*-genes was primarily determined by the variation of non-TNLs.

One of the striking features of the analyses was the topology difference between TIR and non-TIR trees. The TIR-NBS results detected "species-specific" clades (Fig. 3), suggesting that expansion occurred after divergence. Therefore, these types of NBS genes would likely take charge of recognizing "species-specific" pathogens. However, in the non-TIR NBS tree, most NBS genes within a clade were from at least two or more species, indicating their ancient origins originated prior to a split between species, even before the monocot and dicot divergence (Fig. 4). On the other hand, apparently orthologous genes, which existed in at least three plant genomes, could be detected in no less than 12 clades in the non-TIR NBS phylogeny (Fig. 4). However, orthologous genes were not detected in the TIR-NBS subfamily (Fig. 3). This unusual variation in the non-TIR-NBS genes suggested these genes play some basic defense roles, possibly as a result of regional adaptation to the biotic environment, owing to balancing selection.

Most "species-specific" TIR-NBS genes were concentrated on one chromosome and typically arranged in one or

a small number of genomic clusters, e.g. on chromosome 18 in grapevine and chromosome 19 in poplar (Table 2; Fig. S1 and S2). In particular, 55.7% (54 out of 97) of the TNLs and 57.1% (8 out of 14) of the TNs were localized on grapevine chromosome 18, indicating ongoing tandem duplication for the expansion of this family. These results were consistent with the birth-and-death theory (Michelmore and Meyers 1998). The results also suggested a link between the genome physical organization and the origin of gene structure novelties.

Atypical structures in both non-TIR-NBS and TIR-NBS genes were observed in our study. In the poplar non-TIR-NBS group, an intriguing Zf_BED-NBS clade comprised of 59 members was detected (Fig. S4), which could result from the fusion of an N-terminal BED binding zinc finger domain with the CC-NBS homolog, and subsequently duplicated. Another unusual domain, RPW8, was identified in a few non-TIR-NBS genes in the two woody species and several TIR-NBS genes with atypical domain arrangements, including TNLT, TNLTN, TNLTNL, were also found (Table 1). *R*-genes with similar structure have been characterized in *Arabidopsis* (Meyers et al. 2002) and *Medicago* (Ameline-Torregrosa et al. 2008), which were hypothesized to be a fusion of different genes, e.g. a TNTNL gene was a fusion of a TN and TNL gene (Meyers et al. 2002). These findings suggested that domain co-option frequently occurred in NBS genes.

Co-evolution between TIR-NBS-encoding genes and TX genes

Thirty TIR-X genes, which lacked both NBS and LRR domains, were previously identified in *Arabidopsis* (Meyers et al. 2002). These gene products might be analogues of small TIR-adapter proteins that functioned in mammalian innate immune responses (Meyers et al. 2002). In our study, 27, 108 and 1 TXs were identified in grapevine, poplar and rice, respectively (Table 1). Interestingly, approximately 70.4 and 64.5% TXs in grapevine and poplar were found residing in complex clusters, mixed with TN or TNL genes (Fig. S1 and S2). Phylogenetic analysis of the TIR domain showed a pattern similar to that of the TIR-NBS genes, which represented a species-specific expansion. Furthermore, all TXs were found to reside in mixed phylogenetic clades with TNs or TNLs, except for one unique TX subclade, which included 7 TXs from all four-plant genomes (Fig. S6). The phylogenetic relationship and chromosomal position of TXs to TNs and TNLs suggested TXs were derived from TNs or TNLs and co-evolved. An explanation for the mixed clusters, phylogenetic clades or conserved duplications among TX, TN and TNL genes remains equivocal. However, natural selection may have favored these alleles as functional units.

In the particular TX subclade (Fig. S6), one, one, two and three homologs were detected in rice, grapevine, *Arabidopsis* and poplar, respectively. The average identity between the TX gene in rice (*BAF21932*) and the six dicot TXs was 55.6% (ranging from 53.7 to 58.4%), suggesting a highly conserved gene indicated by identity values and similar copy numbers. It was previously hypothesized that TIR encoding genes were lost in grass genomes (Bai et al. 2002). However, here we detected one TIR-X gene (*E*-value $<10^{-11}$ from Pfam) in rice. It remains unclear why grass genomes have a greatly reduced set of TIR encoding genes. This could be a lack of TIR encoding gene amplification or gene loss.

Conclusion

To date, the majority of pathogen resistance genes characterized in plants have provided monogenic dominant resistance. Furthermore, most of these genes were allied with the NBS-LRR class of resistance genes and tremendous progress has been made in understanding the evolution and molecular mechanisms of plant NBS-LRR gene-based resistance. Compared with annual plants, the long-generation time of woody species prevents a comparable rate of evolutionary change relative to microbial or insect pests. However, in this study, we found a significant excess of recent duplications and a higher homologous recombination rate in grapevine and poplar, which could compensate for life history traits and generate novel resistance profiles. Furthermore, an intriguing Zf_BED-NBS-encoding phylogenetic clade, including 59 members, was detected in poplar, indicating that some unique *R*-genes could exist in some species. In addition, TIR-NBS-encoding genes revealed an extensive species-specific expansion. Most of the phylogenetic clades in non-TIR-NBS-encoding genes were mixed with genes from at least three species, suggesting different evolutionary patterns between these two gene types. Finally, a significant negative correlation between K_a/K_s and K_s with NBS duplicates was observed in grapevine, poplar, *Arabidopsis*, and rice, indicating that positive selection or relaxation of negative selection occurred soon after duplication.

Acknowledgments This work was supported by the National Natural Science Foundation of China (30570987 and 30470122), Pre-program for NBRPC (2005CCA02100) and 111 project to D. T., or J-Q. C.

References

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of NBS-LRR genes in the model plant medicago truncatula. Plant Physiol 146:5–21

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res 12:1871–1884

Bailey TL, Elkan C (2005) The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3:21–29

Bent AF, Mackey D (2007) Elicitors, effectors, and *R*-genes: the new paradigm and a lifetime supply of questions. Annu Rev Phytopathol 45:399–436

Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol 54:548–562

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833

Ding J, Araki H, Wang Q, Zhang P, Yang S, Chen JQ, Tian D (2007a) Highly asymmetric rice genomes. BMC Genomics 8:154

Ding J, Zhang W, Jing Z, Chen JQ, Tian D (2007b) Unique pattern of *R*-gene variation within populations in *Arabidopsis*. Mol Genet Genomics 277:619–629

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2:516–527

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiang H, Wang C, Ping L, Tian D, Yang S (2007) Pattern of LRR nucleotide variation in plant resistance genes. Plant Sci 173:253–261

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet 20:116–122

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7:212

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

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, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. Plant J 32:77–92

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

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

Nobuta K, Ashfield T, Kim S, Innes RW (2005) Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in *Arabidopsis*. Mol Plant Microbe Interact 18:103–109

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf–4/9 locus of tomato. Cell 91:821–832

Rice Chromosomes 11 and 12 Sequencing Consortia (2005) The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biol 3:20

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Shen J, Araki H, Chen L, Chen JQ, Tian D (2006) Unique evolutionary mechanism in *R*-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. Genetics 172:1243–1250

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

Xiao S, Ellwood S, Calis O, Patrick E, Li T et al (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291:118–120

Xiao S, Emerson B, Ratanasut K, Patrick E, O'Neill C et al (2004) Origin and maintenance of a broad-spectrum disease resistance locus in *Arabidopsis*. Mol Biol Evol 21:1661–1672

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

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

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I et al (2006) The genome of black cottonwood, *Populus trichocarpa* (torr. & gray). Science 313:1596–1604

van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol 45:43–72

Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen JQ, Tian D (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol Biol 62:181–193

Yang S, Jiang K, Araki H, Ding J, Yang YH, Tian D (2007) A molecular isolation mechanism associated with high intra-specific diversity in rice. Gene 394:87–95

Yu J, Wang J, Lin W, Li S, Li H et al (2005) The Genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zhang P, Gu Z, Li WH (2003) Different evolutionary patterns between young duplicate genes in the human genome. Genome Biol 4:R56

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS genes. Mol Genet Genomics 271:402–415