Pak. J. Bot., 47(5): 1783-1792, 2015.

RAPID DUPLICATION AND LOSS OF NBS-ENCODING GENES IN EUROSIDS II

SHABANA MEMON[1,2¶], WEINA SI[1¶], LONGJIANG GU[1], SIHAI YANG[1*] AND XIAOHUI ZHANG[1*]

[1]School of life Sciences, Nanjing University, Nanjing, 210093, China,
[2]Lecturer, Department of Plant Breeding and Genetics, Sindh Agriculture University, Tando Jam, Hyderabad, 70060, Pakistan.
[*]Corresponding authors: email: xiaohuizhang@nju.edu.cn or sihaiyang@nju.edu.cn
Tel: +86-25-89686406; Fax: +86-25-83592705
[¶]These authors contributed equally to this work

Abstract

Eurosids basically evolved from the core Eudicots 'Rosids'. The Rosids consist of two large assemblages, Eurosids I (Fabids) and Eurosids II (Malvids), which belong to the largest group of Angiosperms, comprising of >40,000 and ~15,000 species, respectively. Although the evolutionary patterns of the largest class of disease resistance genes consisting of a nucleotide binding site (*NBS*) and leucine-rich repeats (*LRRs*) have been studied in many species, systemic research of *NBS*-encoding genes has not been performed in different orders of Eurosids II. Here, five Eurosids II species, *Gossypium raimondii*, *Theobroma cacao*, *Carica papaya*, *Citrus clementina*, and *Arabidopsis thaliana*, distributing in three orders, were used to gain insights into the evolutionary patterns of the *NBS*-encoding genes. Our data showed that frequent copy number variations of *NBS*-encoding genes were found among these species. Phylogenetic tree analysis and the numbers of the *NBS*-encoding genes in the common ancestor of these species showed that species-specific NBS clades, including multi-copy and single copy numbers are dominant among these genes. However, not a single clade was found with only five copies, which come from all of the five species, respectively, suggesting rapid turn-over with birth and death of the NBS-encoding genes among Eurosids II species. In addition, a strong positive correlation was observed between the Toll/interleukin receptor （TIR） type NBS-encoding genes and species-specific genes, indicating rapid gene loss and duplication. Whereas, non-TIR type NBS-encoding genes in these five species showed two distinct evolutionary patterns.

Key words: *NBS*-encoding genes; Evolution; Eurosids II; Fast gene birth and death; Distinct evolutionary patterns.

Introduction

Plant species are constantly exposed to various kinds of pathogens. As a result, plants have evolved various mechanisms to defend invasion. An innate immune systems which called effector-triggered immunity (ETI), wherein the resistant (*R*) genes play a central role in the recognition of pathogen effectors, shows critical importance in disease and pest resistance (De Young & Innes, 2006; Hofberger *et al*., 2014; Tang *et al.*, 2010; Zhang *et al*., 2014). The nucleotide binding site-leucine-rich repeat (*NBS-LRR*) genes form the largest family of *R*-genes. Moreover, according to the N-terminal of the NBS-encoding proteins, the *NBS*-encoding genes can be subdivided into two major subfamilies: one with the Toll/interleukin receptor homolog (TIR) region and another without TIR (non-TIR) region (Meyers *et al.*, 2005). Despite that *TIR-NBS* subclass play an important role in dicotyledons and can be found in the gymnosperms, rare such subclass were discovered in monocots (Tarr & Alexander, 2009).

In plants, the *NBS*-encoding genes recognize avirulence (*Avr*)-gene-specified pathogen molecules by their LRR domains and initiate defense by signal transduction through their conserved NBS domains. These genes have the ability to evolve new '*R*' gene specificities rapidly in response to the rapid variations of the pathogen (Belkhadir *et al*., 2004; Plocik *et al.*, 2004). A total of 174 *R*-genes have been discovered in *Arabidopsis thaliana* (Yang *et al*., 2008), 483 in *Populus trichocarpa* (Yu *et al*., 2014), 545 in *Vitis vinifera* (Yu *et al.*, 2014), 201 in *brassica rapa* (Hofberger *et al*., 2014), 459 in *Citrus sinensis* (Hofberger *et al*., 2014), 245 in *Sorghum bicolor* (Li *et al*., 2010) and 508 *R*-genes in *Oryza sativa* (Li *et al*., 2010). Large scale gene losses and duplications among *NBS*-encoding genes of different species have been observed, which is in concurrence with the assumption that these genes originated from the same ancestor (Li *et al.*, 2010; Meyers *et al*., 2005; Noir *et al*., 2001; Plocik *et al.*, 2004; Yang *et al*., 2008; Yu *et al*., 2014) *NBS*-encoding genes are found to be in clusters in the plant genomes and are also present at specific loci conferring gene duplication and amplification events (Marone *et al.*, 2013). Such *R*-gene family is privileged by tandem repeats due to different birth and death rates (Cannon *et al*., 2004) which is consistent with an origin that is produced through duplication and maintains evolution (Ronald, 1998). Additionally, a significant selection force could be detected on different resistance genes. For instance, in rice blast resistance locus of Pi2/9, positive selection appeared to be the main force promoting divergence , especially in the LRR region (Zhou *et al.*, 2007). Over half of detected NBS-LRR genes in *Arabidopsis* were found under positive selection (Mondragon-Palomino *et al.*, 2002). It might be possible that, *NBS*-encoding genes stand stronger to natural selection due to their interaction with rapid involvement with pathogens. Eudicots, which are characterized by the production of tricolpate or tricolpate-derived pollen grains, comprise 75% of the extant angiosperm with about 200,000 species (Magallon *et al*., 1999). The Eudicots/Tricolpates are classified into two groups: 'Basal Eudicots' and 'Core Eudicots'. Within the core Eudicots two largest groups are present the 'Rosids' and 'Asterids'. The Rosids, a large monophylogenetic clade consists of two main assemblages, Eurosids I (Fabids) and Eurosids II (Malvids) (Forest & Chase, 2009). Eurosids I contribute over 40,000 species, whereas, Eurosids II consist of ~15,000 species (Chase and Reveal, 2009). Although the evolutionary characters of *R*-genes have been well studied in many species (Hofberger *et al*., 2014; Li *et al*., 2010; Noir *et al.*, 2001; Yang *et al.*,

2008; Yu *et al*., 2014), systematic analysis has not been performed to investigate the relationships and evolutionary process of NBS genes among species in different orders of Eurosids II (Zhang *et al*., 2014). In the present study, we selected five important Eurosid II species with whole sequenced genomes, including *Gossypium raimondii*, *Theobroma cacao*, *Carica papaya*, *Citrus clementina*, and *Arabidopsis thaliana*. *T. cacao* and *A. thaliana* belong to the order Brassicales, *G. raimondii* and *C. papaya* are included in the order Malvales, and *C. clementina* in Sapindales (Chase & Reveal, 2009; Forest & Chase, 2009). *A. thaliana* is the model organism of choice in plant biology and genetics which has been successfully used to study the interaction between plants and pathogens. The other four species studied are all economically important crops grown on a commercial scale. Cotton plants have a vital importance in industrial agriculture and trade of many tropical and subtropical countries. The species selected for our study was *G. raimondii*, which is an important putative progenitor of the economically important fiber-producing cotton species. *T. cacao* (*Theobroma Cacao*) belongs to the genus *Theobroma,* which is derived from the Latin word meaning "food of the gods" and its seeds (cocoa beans) are widely used to make cocoa mass, cocoa powder, and chocolate. *C. papaya* and the clementine mandarin (*C. clementina*) are commonly known for their nutritional values in human health. Although, these plants are cultivated worldwide and have a substantial yield annually. They are susceptible to attack by various pests and diseases caused by microbes, such as cocoa capsids and black pod (*T. cacao*), bollworm and alternaria leaf spot (*G. raimondii*), papaya ringspot virus (*C. papaya*) and *Citrus exocortis* and *Citrus cachexia* (*C. clementina*). Breeding of resistant cultivars is an economically important measure to control these numerous diseases. However, few functional resistance genes have been identified in these crops.

In the present study, we revealed the orientation of *NBS* genes, classified into TIR and non-TIR subfamilies, phylogenetically inferred the set of gene families and investigated the gene birth and death rate during their evolutionary process of the five core species of Eurosids II. The fluctuations in gene loss and expansion were observed in the five species by understanding the mechanism of their occurrence. Distinct evolutionary patterns were also discovered between TIR and non-TIR NBS genes. According to this systemic analysis in different orders of Eurosids II, useful information which can promote the genomic research of resistant genes in breeding was concluded.

Materials and Methods

Prediction and classification of NBS genes: For identifying the *NBS*-encoding genes, we screened out the complete genomic sequences and annotations of the five species mentioned above (http://www.phytozome.net). We directly used the estimated proteins for predicting the *NBS*-encoding genes. Firstly, we selected all of the candidate genes which presented NBS-ARC domains from Pfam results version 22.0 (*E* value cut-off of 10^{-4}). All these candidate genes were regarded as NBS-encoding genes. Secondly, all of the NBS-encoding genes were utilized to subsequently evaluate the TIR, CC and LRR motifs through Pfam database (*E* value cut-off of 10^{-4}), SMART protein motif analysis and COILS with a threshold of 0.9 were used to detect CC domains (Lupas *et al.*, 1991).

Sequence alignment and phylogenetic analysis: Although extremely diversity occurred in the gene sequences, NBS domains contribute more to the conserved motif using to distinguish *NBS*-encoding genes from other genes (Ellis *et al.*, 2000). Generally, for *NBS-LRR* genes, the LRR region following the NBS have high variability and are not included for phylogenetic construction (Chen *et al.*, 2010; Li *et al*., 2010; Wan *et al*., 2013; Yang *et al.*, 2008; Yue *et al*., 2012). Therefore, only the conserved NBS regions were used to generate multiple sequence alignments and phylogenetic tree constructions. To perform phylogenetic analyses, all NBS-containing proteins were trimmed to extract the NBS domain as revealed by Pfam and multiple alignments of these amino acid sequences were performed by ClustalW with a default options (Thompson *et al.*, 1994). The aligned amino acid sequences were transferred to nucleotide sequences again and used to construct a phylogenetic tree using MEGA 5.0 (Tamura *et al*., 2007), with neighbor-joining (NJ) method. A Kimura two-parameter model and the internal node stability were explored with 1000 replicates. To find a proper criterion for classifying the gene families, two steps were adopted. In the 1st step, the phylogenetic tree was divided into different clades based on the bootstrap ≥60% and tried to find clades containing alleles with single copy from the five surveyed species. The nucleotides divergence was then estimated. According to the assumption that these core species evolved from an ancestor, having enhance nucleotide divergence was considered as the maximum divergence among the NBS-encoding genes of these five core species. We then classified the phylogenetic tree into different gene families according to the criterion that bootstrap values applied as ≥ 60% and the maximum divergence. Nucleotides divergence among members in gene families were estimated with the Jukes and Cantor correction (Lynch and Crease, 1990) using DnaSP v4.0 (Rozas *et al.*, 2003) Clades consisted of members from more than one species are defined as multi-species gene families, while clades where individual species members were detected were termed as species-specific gene families.

The ratio of the number of non-synonymous substitutions per non-synonymous site (*Ka*) to the number of synonymous substitutions per synonymous site (*Ks*) is termed *Ka/Ks* (Hurst, 2002). The ratio of *Ka/Ks* >1 in protein-coding sequences can be regarded as the evidence of positive selection (Yang & Bielawski, 2000). In order to detect the selection force in different type gene families, the ratio of *Ka/Ks* was calculated among the members in species-specific (paralogs) and multi-species (homologs) gene families, respectively. Protein sequences of members in different gene families were aligned by ClustalW (Thompson *et al.*, 1994). And the results were used to guide the alignment of nucleotide coding sequences by MEGA 5 (Tamura *et al.*, 2007). As illustrated by previous studies (Li *et al*., 2010), the non-synonymous (Ka) and synonymous (Ks) nucleotide substitutions were calculated based on the Nei-Gojobori method with Jukes–Cantor correction (1986) by Dnasp v4.0 (Nei & Gojobori, 1986; Rozas *et al.*, 2003).

Estimation of gene birth and gene death rate in five species: According to the assumption that all members in NBS gene families originated from a common ancestor, the gene birth and gene death ratio of NBS genes in the five species was estimated as described by Li *et al.* (2010). First, we designated the number of clades we identified in the phylogenetic tree as the number of NBS-encoding genes in the common ancestor. Then, a species tree was constructed and the time that these species split from each other were also noted according to previous studies (Chase and Reveal, 2009; Forest & Chase, 2009). Gene numbers in each node was calculated backwards from the latest nodes. And gene numbers of birth and death were calculated based on the gene families and single gene clades we identified. For example, when compared *C. papaya* and *A. thaliana*, we found x gene clades including genes of *C. papaya*, while y *NBS*-encoding genes were identified in *C. papaya*. Thus we regarded y-x as the gene expansion number in *C. papaya* compared with its ancestor common with *A. thaliana*. On the other hand, there were z gene clades including *A. thaliana*, but lack of *C. papaya* genes. It was deduced z as the gene death number in *C.papaya*. While, the rate of gene birth and death was calculated using gene death or birth numbers divided the time they split.

Results

Identification of *NBS*-encoding genes in five Eurosids II species: 420, 51, 283, 300, and 167 *NBS*-encoding genes were identified from *C. clementina*, *C. papaya*, *T. cacao*, *G. raimondii* and *A. thaliana*, respectively (Table 1). *C. clementina* contributed the largest number of NBS-encoding genes among the five species and also accounted for the highest proportion of all the predicted genes (1.71%). In contrast, *C. papaya* incorporated the least *NBS* genes with only 51, accounting for 0.19% of the total number of estimated genes in the whole genome (Table 1), which is consistent with previous reports (Wan *et al*., 2013). The other three species had approximate proportions of *NBS*-encoding genes on the genome scale, 0.6.-1.00% (Table 1).

The NBS-encoding genes were further classified on the basis of their domains such as CC, TIR, and LRR domains. Apparently, most *NBS*-encoding genes (72.5-92.0%) contained the LRR domains. Overall, 375, 37, 253, 276, and 144 *NBS-LRR* genes comprising of both *TIR* and non-*TIR-NBS LRR* were identified in *C. clementina*, *C. papaya*, *T. cacao*, *G. raimondii*, and *A. thaliana*, respectively (Table 1). In case of genome size *C. clementina* and *C. papaya* showed significant difference in numbers of identified *NBS-LRR* genes. The other three species accounted for little variation in the proportions of *NBS-LRR* genes to all predicted genes, which were 0.53%, 0.74%, and 0.86% in *A. thaliana*, *G. raimondii*, and *T. cacao* (Table 1). Since TIR and non-TIR are the two big subfamilies of the *NBS*-encoding genes; the *NBS*-encoding genes were also identified into two groups, *TIR-NBS-LRR* (*TNL*) and non-*TIR-NBS-LRR* (non-*TNL*). Interestingly, the two subclasses were distributed unevenly in the five surveyed species (Table 1). In general, more non-*TNL* than *TNL* genes were detected except in Arabidopsis. About 62.5% *NBS-LRR* genes in Arabidopsis were *TNL* genes, whereas only 7.1-31.2% genes encoded TNL proteins in the other four species (Table 1). Moreover, most of the non-*TNL* genes belonged to the CNL subclass and are prevalent in monocots and dicots. We found *TIR-CC-NBS-LRR* genes in *C. clementina* (11), *G.raimondii* (4), *C. papaya* (2) and *A. thaliana* (9), except in *T. cacao*. These *NBS-LRR* genes encoded TIR and CC domains simultaneously, indicating that sequence exchange events or gene infusion/fission events might have happened before.

Table 1. Identification of *NBS*-encoding genes among five species.

Predicted protein domains	Letter CODE	*Citrus clementina*	*Gossypium raimondii*	*Theobroma cacao*	*Carica papaya*	*Arabidopsis thaliana*
NBS-LRR type genes						
TIR-NBS-LRR types genes						
TIR-NBS-LRR	TNL	83	22	13	4	73
TIR-CC-NBS-LRR	TCNL	11	4	0	2	9
X-NBS-LRR	XNL	23	5	5	3	6
X-CC-NBS-LRR	XCNL	0	2	0	0	2
TOTAL		117	33	18	9	90
NON-TIR-NBS-LRR types genes						
CC-NBS-LRR	CNL	185	161	174	11	37
X-NBS-LRR	XNL	73	82	61	17	17
TOTAL		258	243	235	28	54
TOTAL NBS-LRR genes		375	276	253	37	144
NBS type genes						
TIR-NBS	TN	10	0	2	1	12
X-NBS (TIR type)	XN	0	0	1	1	1
TIR-CC-NBS	TCN	1	0	0	0	3
X-NBS (CC type)	XN	11	8	16	9	6
CC-NBS	CN	23	16	11	3	1
TOTAL		45	24	30	14	23
Total NBS-encoding genes		420	300	283	51	167
Genome size (MB)		367	775	430	372	125
Predicted protein in a genome		24533	37,505	29452	27332	27416

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

Phylogenetic analysis and classification of NBS-encoding gene families: To further evaluate evolutionary and phylogenetic relationships among the *NBS*-encoding genes of the five species, the nucleotides which encoded NBS-domains and generated a phylogenetic tree were selected (Fig. 1, see Methods for details). The genes in the phylogenetic tree were then divided into different clades according to their topologies. The criterion of bootstrap values applied as ≥ 60%, which showed that no clades containing single-copy orthologous genes from the five surveyed species were present. This information was intended for use in estimating the nucleotide divergence among the *NBS*-encoding genes. However, we detected two conserved clades whose members were from all of the five species, but consisted of more than one gene in some species (Fig. 2), suggesting that the two clades might be relatively conserved as compared with other clades among the core species, although few duplicated events have been undergoing in one or two species after the split of these species (Fig. 2).

Subsequently, we calculated the nucleotide divergence (*Dxy*) among the members within these two conserved clades, which ranged from about 11% to 45% among each two members (Fig. 1). According to the assumption that these core species evolved from an ancestor, a rough estimate of 50% was made to evaluate the maximum divergence (*Dxy*) among the *NBS*-encoding genes of these five core species. Therefore, the phylogenetic tree was further subdivided into gene family clades based on the criterion of bootstrap values applied as ≥ 60% and divergence (π) ≤ 50% (Li *et al*., 2010).

The 55 multi-species gene families included 380 genes (~31.1% of the total *NBS* genes) from 2 to 5 species, ranging from 1 to 31 in each species among each family. Only two gene families contained genes from the five species, and they both belonged to non-TIR gene families (Fig. 2). Moreover, among these 55 multi-species gene families, only one TIR and three non-TIR gene families had genes from the four surveyed genomes, suggesting rapid copy number variations of these genes among these species. Interestingly, approximately 76.4% (42/55) multi-species family clades involved NBS-encoding genes identified only from two of the five genomes. Among these, most (31/42) were cacao-cotton specific clades (Table 1), suggesting a closer phylogenetic relationship between cacao and cotton. Meanwhile, all the cacao-cotton specific clades had genes encoding non-TNL proteins, except one TNL-type family. The remaining nine multi-species gene families were identified as three-species gene families, including five cacao cotton clementine families, two cacao cotton papaya families, one each of cacao cotton arabidopsis, and cacoa clementine papaya families (Table 2). Our results suggested that the genes in clementine had lower nucleotide diversity in cacao, cotton or papaya than in Arabidopsis indicating that clementine has a closer phylogenetic relationship with cacao and cotton. Other noteworthy studies were also found that compared with family Asteracae, NBS genes in Arabidopsis became distinct revealing gene duplication (Plocik *et al*., 2004).

Rapid birth and death of the NBS-encoding proteins among core-species: Evolutionary variation was attributed to the rate of gene duplication among the five core species. Different studies highlight that gene families arise through tandem duplication and tremendous rate of birth and death clusters (Cannon *et al*., 2004). Previous studies have shown that generally speciation occurs after gene duplication (Chandan & Indra, 2014; Nei & Nozawa, 2011). We obtained 367 gene clades as determined by bootstrap values ≥ 60% and divergence (π) ≤ 50%. However, not all of the gene clades involved *NBS*-encoding genes from the five species. Only 134 of 367 clades contained genes from *T. cacao*, whereas 110 from *G. raimondii*, 122 from *C. clementina*, 33 from *C. papaya*, and 45 from *A. thaliana* were detected. Under the assumption that the five species evolved from a common ancestor, 367 could be the minimal NBS-encoding gene number acquisition within their ancestor. Hence, the dynamic numbers of the *NBS*-encoding genes in the five species was estimated (Fig. 3). A number of gene birth and gene death events might have occurred among these five surveyed genomes containing *NBS*-encoding *R*-genes, suggesting that these genes experience rapid birth-and-death evolution during speciation.

In general, we estimated that the birth rate was higher than the death rate in the four genomes, except in *C. papaya*. *C. clementina* is estimated to have diverged from Malvales 85 million years ago (Fig. 4). Citrus belonging to the order Sapindales is supposed to have a closer relationship with Brassicales and Malvidae (Xu *et al*., 2013). Interestingly, in our results, a higher birth rate rather than death rate contributed to the split with the ancestor of cacao and cotton, with 3.51 and 2.16 copies per mya respectively. Thus, the largest number of *NBS*-encoding genes in *C. clementina* could be attributed to a high birth rate.

Comparative genomics studies showed that papaya is a useful out-group that evolved from WGD in the Arabidopsis lineage, and the ancestors of papaya and arabidopsis split about 72 million years ago (mya) (Ming *et al*., 2008). The present study suggests that *C. papaya* and *A. thaliana* underwent significant gene losses after their split with other three species. However, *C. papaya* experienced less gene birth (0.25 per mya) and more gene death (0.58 copies per mya) in contrast to higher gene birth (1.69 per mya) and less gene death (0.42 per mya) in *A. thaliana*. Thus, only 51 NBS-encoding *R*-genes were retained in the *C. papaya* genome (Fig. 3). Paleohistorical studies showed that the eudicot genomes (cacao, papaya and arabdopsis) were derived from a common ancestor and passed through paleohexaploidization events following numerous lineage specific whole genome duplications (WGDs) (Zhang *et al*., 2014). This variation in gene expansion and contraction helped to explore maximum divergence among the five species from their common ancestor; thus, providing new, rapid evolution of *R*-genes in the Eurosid II species.

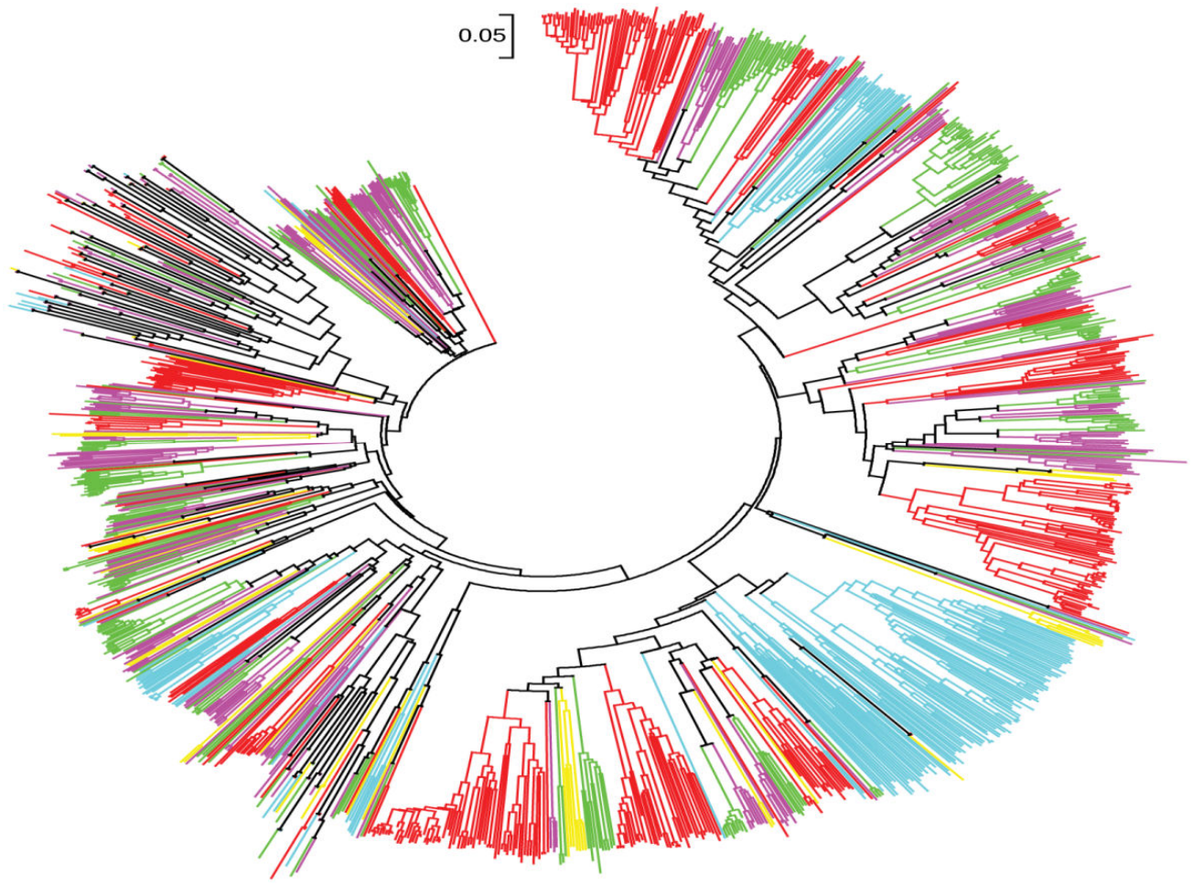

Fig. 1. Phylogentic tree consisting of different clades of five species.
The phylogenetic tree was constructed using MEGA 5.0, with neighbor-joining (NJ) method and was explored with 1000 replicates. The clades consisted of TNLS and NTNLS NBS-encoding *R*-genes, each contributing five species with different clade colors. *C.papaya* resembles yellow clade colour, *C.clementina* red colour, *G.raimondii* green colour, *T.cacao* pink colour and *A.thaliana* light blue colour respectively.

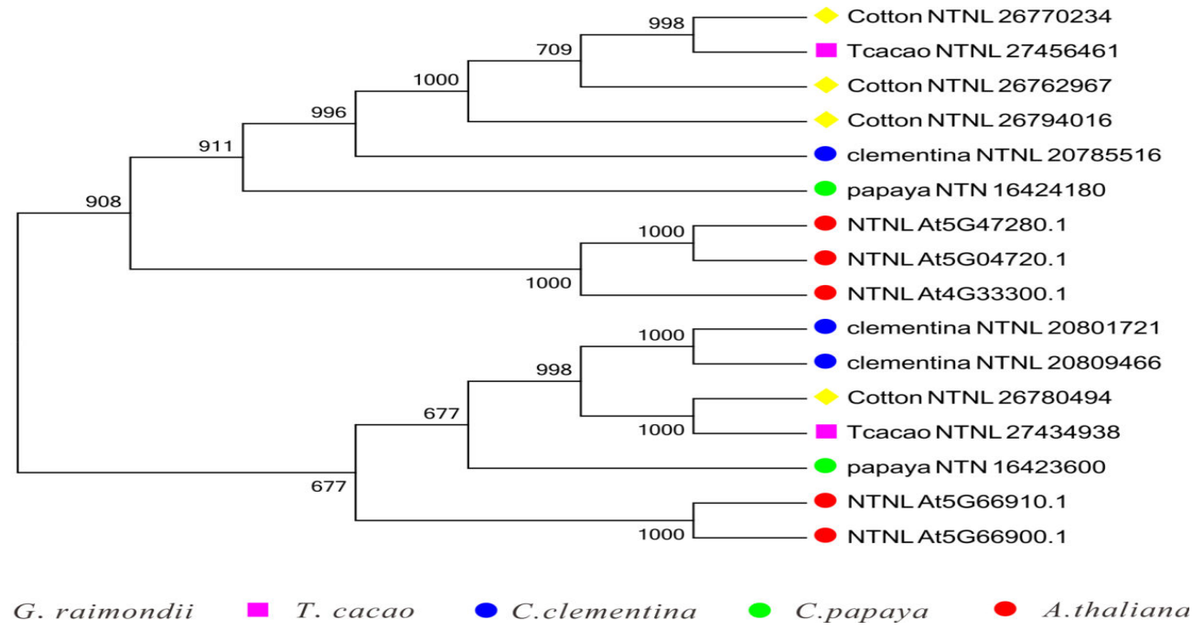

Fig. 2. Conserved clade with non-TIR type *NBS*-encoding genes have different evolutionary pattern.
The Part of tree representing evolutionary pattern in two conserved clades including the five core species which are all NTNLs (non-TIR *NBS-LRR* encoding genes).The shapes and colours representing as follows:

Table 2. Gene loss and duplication patterns among five species.

Multi-species clades	Family member	Average gene numbers				
		T. cacao	*G. ramondii*	*C. clementina*	*C. papaya*	*A. thaliana*
TO	31	3.97	3.19	-	-	-
TL	3	1.33	-	1.33	-	-
TP	2	1.00	-	-	3.00	-
TA	2	1.50	-	-	-	1.00
LP	1	-	-	25.00	1.00	-
LO	1	-	1.00	2.00	-	-
LA	1	-	-	5.00	-	3.00
PA	1	-	-	-	1.00	1.00
TOA	1	1.00	1.00	-	-	1.00
TOL	5	2.40	2.20	1.20	-	-
TOP	2	4.50	8.00	-	2.00	-
TLP	1	1.00	-	8.00	1.00	-
TOLA	2	1.00	2.00	1.00	-	1.50
TOLPA	2	1.00	2.00	1.50	1.00	2.50
sum	55	3.12	3.09	3.44	1.67	1.67
Species-specific clades						
T	80	1.55	-	-	-	-
O	66	-	2.48	-	-	-
L	106	-	-	3.44	-	-
P	24	-	-	-	1.5	-
A	36	-	-	-	-	4.22
sum	312	1.55	2.48	3.44	1.50	4.22
Total	367	2.16	2.73	3.44	1.55	3.71

T = *T. cacao*, O = *G. raimondii*, L= *C. clementina*, P=*C. papaya*, A=*A. thaliana*. In multi-species families different combinations of abbreviations are used to represent the *R*-genes are shared with different species in a family. For instance TO means *T. cacao* and *G.raimondii*, and TOLPA consists the five species in one family

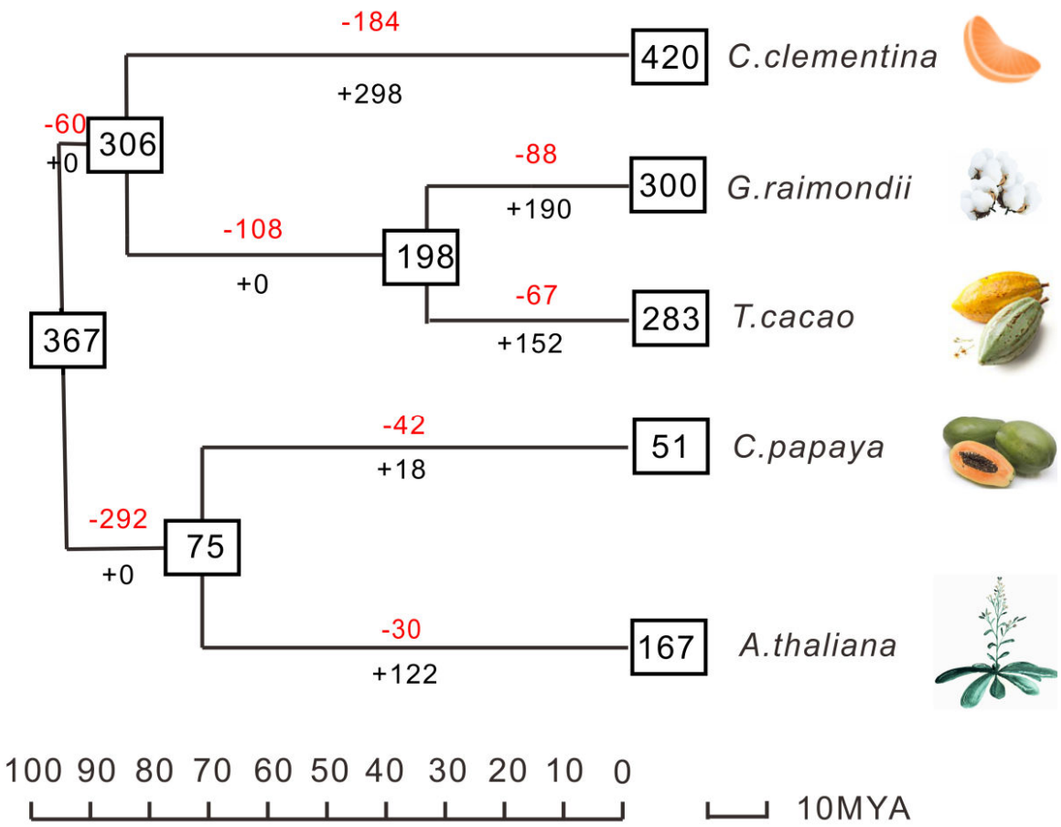

Fig. 3. Gene loss and expansion among five core-eudicots species. The rectangles represent the gene numbers, which were calculated based on the phylogenetic tree and with maximum nucleotide identity with ≤ 50% divergence (π). 367 was the minimum ancestral gene number from which the five species appeared from a common ancestor. The numbers with plus and minus represents the expansion and contraction of the species which conferred that some genes were lost during speciation, on the other hand most of the genes duplicated assuming that they came from a common ancestor.

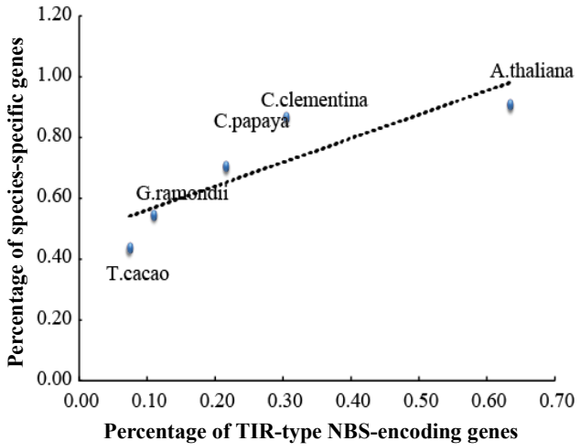

Fig. 4. The correlation analysis of the percentage of *NBS*-encoding genes in species-specific clades and *TIR*-type *NBS*-encoding genes. Relationship between species specific genes and TIR type *NBS*-encoding genes. The Y-axis represents the percentage of species-specific genes and X-axis shows the percentage of *TIR*-type *NBS*-encoding genes.

Higher non-synonymous to synonymous substitution ratios in species-specific NBS-encoding gene families: In the R gene family, non-synonymous to synonymous substitution is significant for distinguishing the evolutionary rates through selection pressure. Previous studies showed that *Ka/Ks*>1 demonstrates positive selection, which leads to diversifying selection, found in different regions of *R*-genes. Whereas, *Ka/Ks* <1 conveys a negative selection and *Ka/Ks*=1 indicates a neutral selection (Li *et al*., 2010; Mondragon-Palpmino *et al*., 2002; Tobias and Guest, 2014; Wan *et al.*, 2012; Yang *et al*., 2008; Zhang *et al*., 2011). Mapping the positively selected amino acid residues involving *NBS*-encoding genes is an important method for exploring gene function and diversifying selection (Mondragon-Palomino *et al*., 2002; Tobias & Guest, 2014). In the present study, mainly two kinds of gene were identified from the phylogenetic tree, namely the multi-species and species-specific gene families. The *Ka/Ks* ratios of CDS were calculated in these gene families to further investigate the differences in their evolution patterns (Table 3). The orthologs indicated strong positive correlation where *Ka/Ks* >1 was not detected in the multi-species gene families, however, one such event was found in the species-specific gene families. On the other hand, the average value of *Ka/Ks* (0.554) in species-specific gene families was larger than that of multi-gene families (0.368) (t-test, p-value=5.227×10^{-10}). We also discovered, eight species-specific gene families contributing more than 0.8 *Ka/Ks* value, suggesting potential positive selection. However, the highest value of *Ka/Ks* of multi-species families was only 0.65 (Table 3). Thus, the above results suggest that the species-specific gene families undergo less negative selection compared to the multi-species gene families.

Discussion

Rapid gene loss and duplication of *NBS*-encoding genes in Eurosids II: Disease resistance genes can broadly detect the presence of pathogens in a number of plant species. It was reported that *R*-genes evolved rapidly and generated variation in gene numbers in different species (Ameline-Torregrosa *et al.*, 2008; Jiang *et al*., 2007; Li *et al*., 2010; Plocik *et al.,* 2004; Yang *et al.*, 2008). Although, *NBS*-encoding genes in the five species putatively originated from a common ancestor, variation in the number of *R*-genes is evident. Our results showed that *C. clementina* had the largest number of *NBS*-encoding genes (420) among the five species, followed by *G. raimondii* (300), *T. cacao* (283), *A. thaliana* (167), and *C. papaya* (51). However, *G. raimondii* was found to contribute with a largest genome size (775Mb). *A. thaliana* had more than three times the *NBS*-encoding genes (167) as compared to *C. papaya* (51) despite having the smallest genome size among the five species (Table 2). This evidence ensures that *NBS*-encoding genes numbers were not corresponding to the genome expansion of the five core species. Furthermore, *NBS*-encoding genes accounted for about 1.71%, 0.80%, 0.96%, 0.19% and 0.61% of all predicted genes in the five core species (Table 2), indicating that the number of these genes was not relative to the genome estimated proteins. This phenomenon has also been observed in grasses (Li *et al.,* 2010).

According to the evolutionary analysis of *NBS*-encoding genes, dominant species-specific gene families and single gene clades were identified, whereas about 80% multi-species gene families were two-species specific gene families. Together with the deduced rate of gene birth and death, our results indicated a rapid birth and death dynamic in NBS-encoding genes. Gene duplication, especially WGD, provides sufficient genomic materials for retention and loss of genes. Loss of redundant *NBS*-encoding genes generated by gene duplication has been reported in many plant species, such as Arabidopsis (Nobuta *et al*., 2005), rice (Yu *et al.*, 2005), grape (Yang *et al*., 2008) and poplar (Yang *et al*., 2008).

C. clementina experienced the rapidest gene duplication and gene loss rates among the five genomes. However, no recent WGD has been detected in *C. clementina* (Xu *et al.*, 2013). Therefore, we can predict that *R-gene* expansion in *C. clementina* was not because genome duplication, but the result of tandem duplication, unequal crossing over, retro-transposition, and duplicated DNA transposition (Zhang, 2003). Previous studies identified orthologs between *G. raimondii* and *T. cacao*, which strictly shared duplicated genes (Argout *et al.*, 2011; Renny-Byfield *et al*., 2014). In our study, the number of *NBS*-encoding genes in *G. raimondii* and *T. cacao* were the closest and most of the multi-gene families were cotton-cacao lineage-specific gene families. In addition, the average copy numbers of gene families in *G. raimondii* (2.76) and *T. cacao* (2.16) are approximately equivalent, in contrast to the recent discovery that *G. raimondii* went through an independent genome duplication after its split with *T. cacao*. However, we found that the average copy number of species-specific gene families in *G. raimondii* (2.48) is almost twice of that in *T. cacao*, which indicate gene expansions occurring in the independent genome duplication events in *G. raimondii*.

Table 3. The distribution of the ratio of non-svnonvmous to synonymous between multi-species and specific-species families.

Multi species family	Ka	Ks	Ka/Ks
		A	
1	1.36	2.84	0.48
2	0.30	0.49	0.63
3	0.23	0.53	0.46
4	0.25	0.48	0.57
5	0.22	0.54	0.40
6	0.32	0.80	0.40
7	0.41	0.72	0.65
8	0.22	0.39	0.64
9	0.34	0.95	0.37
10	0.24	0.55	0.45
11	0.34	0.74	0.57
12	0.15	0.36	0.48
13	0.18	0.78	0.28
14	0.33	0.65	0.50
15	0.29	0.59	0.53
16	0.07	0.36	0.19
17	0.21	1.06	0.21
18	0.07	0.80	0.09
19	0.08	0.37	0.24
20	0.21	0.55	0.42
21	0.24	0.63	0.49
22	0.09	0.37	0.25
23	0.06	0.51	0.11
24	0.34	1.17	0.30
25	0.49	1.05	0.61
26	0.31	0.83	0.42
27	0.16	0.50	0.32
28	0.20	0.46	0.47
29	0.18	0.43	0.48
30	0.13	0.33	0.43
31	0.30	0.86	0.42
32	0.33	0.98	0.34
33	0.13	0.80	0.17
34	0.37	0.99	0.42
35	0.48	1.92	0.25
36	0.41	2.03	0.22
37	0.45	1.63	0.39
38	0.21	0.66	0.47
39	0.30	0.97	0.33
40	0.49	1.13	0.47
41	0.44	1.90	0.27
42	0.45	1.81	0.25
43	0.32	1.64	0.23
44	0.30	1.05	0.30
45	0.31	1.61	0.21
46	0.35	1.73	0.22
47	0.26	1.10	0.29
48	0.40	1.38	0.31
49	0.39	1.13	0.38
50	0.37	1.22	0.41
51	0.26	0.90	0.48
52	0.29	1.22	0.37
53	0.16	0.97	0.17
54	0.24	1.63	0.16
55	0.42	1.69	0.28
Multi species family	**Ka**	**Ks**	**Ka/Ks**
		B	
1	0	0	-
2	0.21	0.49	0.46
3	0.30	0.71	0.45
4	0.16	0.33	0.53
5	0.22	0.42	0.55
6	0.20	0.49	0.51
7	0.33	0.83	0.40
8	0.34	0.75	0.49
9	0.15	0.31	0.52
10	0.07	0.09	0.85
11	0.30	0.74	0.41
12	0.19	0.51	0.39
13	0.12	0.27	0.45
14	0.10	0.14	0.69
15	0.03	0.03	0.76
16	0.16	0.50	0.32
17	0.23	0.71	0.35
18	0.09	0.11	0.79
19	0.06	0.09	0.70
20	0.21	0.47	0.45
21	0.19	0.40	0.48
22	0.18	0.42	0.46
23	0.07	0.09	0.79
24	0.20	0.42	0.49
25	0.29	0.79	0.37
26	0.19	0.30	0.65
27	0.11	0.22	0.51
28	0.11	0.25	0.46
29	0.15	0.33	0.46
30	0.35	0.63	0.59
31	0.28	0.78	0.36
32	0.03	0.04	0.77
33	0.37	1.00	0.37
34	0.12	0.23	0.55
35	0.30	1.06	0.29
36	0.13	0.97	0.13
37	0.43	1.25	0.37
38	0.40	1.32	0.30
39	0.27	0.60	0.45
40	0.52	1.33	0.42
41	0.38	0.92	0.42
42	0.36	0.80	0.47
43	0.37	1.63	0.23
44	0.43	0.65	0.69
45	0.28	0.63	0.46
46	0.52	1.61	0.34
47	0.52	1.18	0.45
48	0.29	0.70	0.45
49	0.04	0.05	0.84
50	0.10	0.12	0.84
51	0.09	0.19	0.49
52	0.41	1.04	0.39
53	0.42	1.27	0.33
54	0.13	0.31	0.40
55	0.55	0.67	0.76
56	0.18	0.20	0.89
57	0.22	0.47	0.52
58	0.09	0.21	0.40
59	0.14	0.14	1.10
60	0.06	0.08	0.67
61	0.24	0.36	0.68
62	0.04	0.06	0.61
63	0.17	0.31	0.53
64	0.12	0.19	0.64
65	0.35	0.69	0.51
66	0.26	0.53	0.57
67	0.27	0.33	0.80
68	0.08	0.10	0.79
69	0.06	0.10	0.69
70	0.08	0.12	0.66
71	0.08	0.08	0.90
72	0.01	0.02	0.57
73	0.08	0.08	0.99
74	0.07	0.11	0.68
75	0.18	0.38	0.50
76	0.15	0.29	0.54
77	0.11	0.16	0.69
78	0.12	0.22	0.58
79	0.31	0.53	0.60
80	0.25	0.49	0.51
81	0.14	0.26	0.58
82	0.12	0.36	0.33
83	0.15	0.34	0.52

In contrast, *C. papaya* and *A. thaliana* had significant *R*-gene loss before their split, which contributed to a lower number of NBS-encoding genes. However, *A. thaliana* underwent two recent WGDs after its split from *C. papaya;* we suppose that it can be regarded as the main reason that *A. thaliana* had more than three times of *NBS*-encoding genes in *C. papaya*. On the other hand, papaya-specific gene families were found to be scarce indicating a lack of recent *R*-gene duplication. Frequent loss and deficient duplications resulted in a low copy number of *R*-genes in papaya. At the time of domestication, the deletion or silencing of duplicated *NBS*-encoding genes might be due to functional redundancy or artificial selection. Different species have been under different selective pressure, which might account for the various numbers and types of retained genes. These findings are similar with Yang *et al.* (2008) in which NBS- encoding genes retain duplicated genes due to tandem repeats.

Different evolutionary patterns between *TIR*-type and non-*TIR* type *R*-genes: *NBS*-encoding proteins could be classified into two types based on the structure of their N-termini: *TIR-NBS*-encoding proteins and non-*TIR-NBS*-encoding proteins. The two subclasses are found to have different evolutionary patterns in many species. For example, their copy number, topology of the phylogenetic tree, selective constraint, and downstream signaling pathways are supposed to be different (Yang *et al*., 2008; Yu *et al.*, 2014; Yue *et al.*, 2012). The present study suggests the abundance of non-*TIR-NBS*-encoding genes in the five Eurosid II species. The phylogenetic tree showed that the *TIR-NBS*-encoding genes and non-*TIR-NBS*-encoding genes were separated into different clades (Fig. 1). Moreover, gene families composed of *TIR* type genes were all species-specific except two (Fig. 1).

Interestingly, we observed significant relationships between *TIR-NBS*-encoding genes and species-specific genes (r^2=0.7535, $p<0.05$) (Fig. 4), indicating TIR-type NBS-encoding genes and species-specific genes might have positive correlation. In molecular genetics, the interaction of speciation is beneficial to several domains with evolutionary computation for calculation of the relationship among them (Kim *et al*., 2011). The average copy numbers of species-specific gene clades of the five species (Table 2) indicated remarkable recent duplications among the five core species. Moreover, relatively stronger positive selection in species-specific clades was detected as compared to multi-species gene clades. Studies have shown that species-specific or lineage-specific gene families contribute mostly to positively selected genes, which are considered as driving forces behind the species (Wang *et al*., 2011). Thus, we can also assume that *TIR*-type *NBS*-encoding genes underwent fast gene loss and expansion, as well as strong positive selection.

On the other hand, only two clades contained genes from all of the five species, which was recognized as relatively ancient clades. Amazingly, all of the members of five core species were non-*TIR* type *NBS*-encoding genes. Many of the multi-species members included non-*TIR* gene clades than *TIR* gene, whereas almost all the clades containing multi-*TIR* type genes were species-specific clades, suggesting different evolutionary patterns in the two types of genes. However, non-*TIR-NBS*-encoding gene clades were species-specific suggesting two different models for the non-*TIR-NBS*-encoding genes. A multi-species gene family usually denotes an older origin and evolves relatively slower and is in sharp contrast with the species-specific genes clades, which indicate fast gene loss and expansion (Yu *et al.*, 2014). Thus, our study suggests that non-*TIR-NBS*-encoding genes have an older origin and might be the ancestor of *TIR*-type *NBS*-encoding genes. Moreover, we found several NBS-encoding genes, which were comprised of both TIR and CC domains indicating that sequence exchange events or gene infusion/fission events might have happened before. However, further evidence is required to explore the complex evolutionary relationship between these two types of NBS-encoding genes.

Acknowledgements

This work was supported by National Natural Science Foundation of China (91331205, J1103512 and J1210026), NSFC of Jiangsu province (BK2011015), Program for Changjiang Scholars and Innovative Research Team in University (IRT1020).

References

Ameline-Torregrosa, C., B.B. Wang, M.S. O'Bleness, S. Deshpande, H. Zhu, B. Roe, N.D. Young and S.B. Cannon. 2008. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. *Plant Physio.,* 146(1): 5-21.

Argout, X., J. Salse, J.M. Aury, M.J.Guiltinan, G. Droc, J. Gouzy, M. Allegre, C. Chaparro, T. Legavre and S.N . Maximova. 2011. The genome of Theobroma cacao. *Nature Genetics.,* 43(2): 101-108.

Belkhadir, Y., R.Subramaniam and J.L. Dangl. 2004. Plant disease resistance protein signaling: NBS–LRR proteins and their partners. *Curr. Opin. Plant Biol.,* 7(4): 391-399.

Cannon, S.B., A. Mitra, A. Baumgarten, N.D. Young and G. May. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.,* 4(1): 10.

Chandan, R. and D. Indra. 2014. Gene duplication: A major force in evolution and bio-diversity. *Inter. J. Biodiversity and Conservation,* 6(1): 41-49.

Chase, M.W. and J.L. Reveal. 2009. A phylogenetic classification of the land plants to accompany APG III. *Bot. J. Linnean Society,* 161(2): 122-127.

Chen, Q., Z. Han, H. Jiang, D. Tian and S. Yang.2010. Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives. *J. Mol. Evol.,* 70(2): 137-148.

DeYoung, B.J. and R.W. Innes. 2006. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology,* 7(12): 1243-1249.

Ellis, J., P. Dodds and T. Pryor. 2000. Structure, function and evolution of plant disease resistance genes. *Cur.Opin. Plant Biol.,* 3(4): 278-284.

Forest, F. and M.W. Chase. 2009. Eurosid II. The timetree of life. Oxford University Press, New York:188-196.

Hofberger, J.A., B. Zhou, H. Tang, J.D. Jones and M.E. Schranz. 2014. A novel approach for multi-domain and multi-gene family identification provides insights into evolutionary dynamics of disease resistance genes in core eudicot plants. BMC genomics.15: 966.

Hurst, L.D. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet*., 18(9): 486-487.

Jiang, H.Y., C.C. Wang, L. Ping, D.C. Tian and S.H. Yang. 2007. Pattern of LRR nucleotide variation in plant resistance genes. *Plant Science*, 173(2): 253-261.

Kim, K.J., J.G. Park and S.-B. Cho. 2011. Correlation analysis and performance evaluation of distance measures for evolutionary neural networks. *J. Intelligent and Fuzzy Systems,* 22(2): 83-92.

Li, J., J. Ding, W. Zhang, Y. Zhang, P. Tang, J.-Q. Chen, D. Tian and S. Yang. 2010. Unique evolutionary pattern of numbers of gramineous NBS–LRR genes. *Mol. Gen. & Geno.,* 283(5): 427-438.

Lupas, A., M. Van Dyke and J. Stock. 1991. Predicting coiled coils from protein sequences. *Science,* 252(5009): 1162-1164.

Lynch, M. and T.J. Crease. 1990. The analysis of population survey data on dna-sequence variation. *Mol. Biol. Evol.,* 7(4): 377-394.

Magallon, S., P.R. Crane and P.S. Herendeen. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Ann. Missouri Bot. Garden.,* 86(2): 297-372.

Marone, D., M.A. Russo, G. Laidò, P. De Vita, R. Papa, A. Blanco, A. Gadaleta, D. Rubiales and A.M. Mastrangelo. 2013. Genetic basis of qualitative and quantitative resistance to powdery mildew in wheat: from consensus regions to candidate genes. *BMC Genomics,* 14(1): 562.

Meyers, B.C., S. Kaushik and R.S. Nandety. 2005. Evolving disease resistance genes. *Curr. Opin. Plant Biol.,* 8(2): 129-134.

Ming, R., S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J.H. Saw, P. Senin, W. Wang, B.V. Ly and K.L. Lewis. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature,* 452(7190): 991-996.

Mondragón-Palomino, M., B.C. Meyers, R.W. Michelmore and B.S. Gaut. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research,* 12(9): 1305-1315.

Nei, M. and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.,* 3(5): 418-426.

Nei, M. and M. Nozawa. 2011. Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. *Gen. Biol. Evol.,* 3: 812-829.

Nobuta, K., T. Ashfield, S. Kim and R.W. Innes. 2005. Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in Arabidopsis. *Mol. Plant-Microbe Interactions*, 18(2): 103-109.

Noir, S., M.-C. Combes, F. Anthony and P. Lashermes. 2001. Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). *Mol. Genetics and Genomics*, 265(4): 654-662.

Plocik, A., J. Layden and R. Kesseli. 2004. Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory. *Mol phyl. Evol.,* 31(1): 153-163.

Renny-Byfield, S., J. P.Gallagher, C.E. Grover, E., Szadkowski, J.T. Page, J.A. Udall, X. Wang, A.H Paterson and J.F. Wendel. 2014. Ancient gene duplicates in Gossypium (cotton) exhibit near-complete expression divergence. *Gen. Biol. Evol.,* 6 (3): 559-571.

Ronald, P.C. 1998. Resistance gene evolution. *Curr.opin. Plant Biol.,* 1(4): 294-298.

Rozas, J., J.C. Sanchez-DelBarrio, X. Messeguer and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics,* 19(18): 2496-2497.

Tamura, K., J. Dudley, M.Nei and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.,* 24(8): 1596-1599.

Tang, P., Y. Zhang, X. Sun, D. Tian, S.Yang and J. Ding.2010. Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. *Plant Science,* 179(4): 399-406.

Tarr, D.E.K. and H.M. Alexander.2009. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Research Notes,* 2(1): 197.

Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22(22): 4673-4680.

Tobias, P.A. and D.I. Guest. 2014. Tree immunity: growing old without antibodies. *Trends in Plant Science,* 19(6): 367-370.

Wan, H., W. Yuan, Q. Ye, R. Wang, M. Ruan, Z. Li, G. Zhou, Z. Yao, J. Zhao and S. Liu. 2012. Analysis of TIR-and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics,* 13(1): 502.

Wan, H.J., W. Yuan, K.L. Bo, J. Shen, X. Pang and J.F . Chen. 2013. Genome-wide analysis of NBS-encoding disease resistance genes in Cucumis sativus and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *Bmc Genomics,* 14.

Wang, D., F. Liu, L. Wang, S. Huang and J. Yu. 2011. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biol. Direct.,* 6: 13.

Xu, Q., L.L.Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W. B. Jiao, B.H. Hao and M.P. Lyon. 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics.,* 45(1): 59-66.

Yang, S., X. Zhang, J.X. Yue, D. Tian and J.Q Chen. 2008. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Gen. Geno.,* 280(3): 187-198.

Yang, Z.H. and J.P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.,* 15(6): 496-503.

Yu, J., S. Tehrim, F. Zhang, C. Tong, J. Huang, X. Cheng, C. Dong, Y. Zhou, R. Qin and W. Hua. 2014. Genome-wide comparative analysis of NBS-encoding genes between Brassica species and Arabidopsis thaliana. *BMC Genomics,* 15(1): 3.

Yu, J., J. Wang, W. Lin, S. Li, H. Li, J. Zhou, P. Ni, W.Dong, S. Hu and C. Zeng. 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biology*, 3(2): 38.

Yue, J.X., B.C. Meyers, J.Q.Chen, D. Tian and S. Yang. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phyt.,*193(4): 1049-1063.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.,*18(6): 292-298.

Zhang, R., F. Murat, C. Pont, T. Langin and J. Salse. 2014. Paleo-evolutionary plasticity of plant disease resistance genes. *BMC Genomics,* 15(1): 187.

Zhang, X., Y. Feng, H. Cheng, D.Tian, S. Yang and J.Q. Chen. 2011. Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. *Mol. Gen. Geno.,* 285(1): 79-90.

Zhou, B., M. Dolan, H. Sakai and G.L. Wang. 2007. The genomic dynamics and evolutionary mechanism of the Pi2/9 locus in rice. *Mol. Plant Microbe Inter.,* 20(1): 63-71.

(Received for publication 6 November 2014)