Genomics 92 (2008) 246–253

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genomewide comparative phylogenetic and molecular evolutionary analysis of tubby-like protein family in *Arabidopsis*, rice, and poplar

Zefeng Yang [1], Yong Zhou [1], Xuefeng Wang, Shiliang Gu, Jianmin Yu, Guohua Liang, Changjie Yan, Chenwu Xu *

Jiangsu Provincial Key Laboratory of Crop Genetics and Physiology, Key Laboratory of Plant Functional Genomics of Ministry of Education, Yangzhou University, Yangzhou, 225009, China

ARTICLE INFO

Article history:
Received 7 December 2007
Accepted 9 June 2008
Available online 23 July 2008

Keywords:
Tubby-like protein gene family
Species-specific expansion
Segmental duplication
Co-evolution

ABSTRACT

Tubby-like proteins, which are characterized by a highly conserved tubby domain, play an important role in the maintenance and function of neuronal cells during postdifferentiation and development in mammals. In additional to the tubby domain, most tubby-like proteins in plants also possess an F-box domain. Plants also appear to harbor a large number of *TLP* genes. To gain insight into how *TLP* genes evolved in plants, we conducted a comparative phylogenetic and molecular evolutionary analysis of the tubby-like protein gene family in *Arabidopsis*, rice, and poplar. Genomewide screening identified 11 *TLP* genes in *Arabidopsis*, 14 in rice, and 11 in poplar. Phylogenetic trees, domain organizations, and intron/exon structures classified this family into three subfamilies and indicated that species-specific expansion contributed to the evolution of this family in plants. We determined that in rice and poplar, the tubby-like protein family had expanded mainly through segmental duplication events. Tissue-specific expression analysis indicated that functional diversification of the duplicated *TLP* genes was a major feature of long-term evolution. Our results also demonstrated that the tubby and F-box domains had co-evolved during the evolution of proteins containing both domains.

In mammals, *TLP* (tubby-like protein) genes play important roles in maintenance and function of neuronal cells during postdifferentiation and development [1]. Tubby, the base for tubby-like proteins, was first identified from obese mice through positional cloning [2,3]. In addition to *TUBBY*, three other members (*TULP1, TULP2,* and *TULP3*) of this gene family that encoded tubby-like proteins have also been identified in humans and mice [4]. Tubby-like proteins in animals are characterized by a highly conserved domain of about 270 amino acids, the tubby (TUB) domain, which is located at the C terminus, but their N-terminal sequences are quite divergent [4]. Until now, tubby-like proteins had been found in many multicellular organisms from both the plant and animal kingdoms. In addition to TUB domains at the C terminus, most tubby-like proteins in plants also contain highly conserved F-box domains [5]. The C terminus of F-box proteins generally contains one or several highly variable protein-protein interaction domains, such as the Leu-rich repeat (LRR), kelch repeat, tetratricopeptide repeat (TPR), and WD40 repeat [6]. The characteristics of the highly conserved TUB domain in different species demonstrated that these proteins must have fundamental biological functions in multicellular organisms.

Comparatively little research, however, has been done in plants. In the model plant *Arabidopsis*, 11 members of this family had been identified, named *AtTLP1* to *AtLP11* [5]. Among them, *AtTLP1, AtTLP2, AtTL3, AtTLP6, AtTLP7, AtTLP9, AtTLP10,* and *AtTLP11* were expressed ubiquitously in all the organs tested, but expression of *AtTLP5* and *AtTLP8* exhibited dramatic organ specificity. Interaction between AtTLP9 and ASK1 has also been confirmed. And transgenic plants overexpressing *AtTLP9* were shown to be hypersensitive to ABA, suggesting that *AtTLP9* may participate in the ABA signaling pathway. In addition to *Arabidopsis*, *TLP* genes have been demonstrated by homology searches to be present in other plants, such as *Lemna paucicostata* [5], *Oryza sativa* [5], *Cicer arietinum* [5], and *Zea mays* [5,7].

It is well known that gene duplication events are important to gene family evolution, which can occur via three major mechanisms: segmental duplication, tandem duplication, and transposition events such as retroposition and replicative transposition [8]. Among these, tandem and segmental duplication events contribute mostly to the generation of new members in nuclear gene families. Cannon et al. [9] analyzed 50 gene families in *Arabidopsis* and reported that tandem duplications were most prominent in some gene families, whereas segmental duplications occurred more frequently in others. Plants, in particular, appear to harbor a large number of *TLP* genes [1]. Establishment of the complete genomic sequences of *Arabidopsis* [10], rice [11–13],

---

Abbreviations: ABA, abscisic acid; EST, expressed sequence tag; ME, minimal evolution; MP, maximum parsimony; NJ, neighbor-joining; OC, orthologous cluster; RT-PCR, reverse transcription polymerase chain reaction; TLP, tubby-like protein; TUB domain, tubby domain.

* Corresponding author. Fax: +86 514 87996817.
E-mail address: qtls@yzu.edu.cn (C. Xu).

[1] These authors contributed equally to this work.

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.06.001

and poplar [14] gave us the opportunity to learn all the *TLP* genes in these three plant species and to investigate the expansion patterns of this family. We identified 11, 14, and 11 *TLP* genes in *Arabidopsis*, rice, and poplar, respectively. A phylogenetic tree was constructed to evaluate the evolutionary relationships of *TLP* genes in the three plant species. We also examined the chromosomal distribution of *TLP* genes to explore potential mechanisms leading to their species-specific expansion in plants. Co-evolutionary analysis of F-box and TUB domains was performed. RT-PCR and in silico data analysis were used to examine tissue-specific expression patterns and functional diversification of paralogous *TLP* genes. To examine the driving force for duplicated genes, we performed nonsynonymous and synonymous rate (K_a and K_s) analyses of the paralogous genes. Our systematic analysis provides a solid foundation for further functional dissection of *TLP* genes in plants.

Results

Collection of TLP genes in Arabidopsis, rice, and poplar

After careful survey of three plant genomes, 11 members of the TLP family in *Arabidopsis*, 14 in rice, and 11 in poplar were identified. The nucleotides, CDS, and protein sequences of *TLP* genes in *Arabidopsis* and rice were downloaded from the TIGR database (Supplementary Table 1), whereas those in poplar were downloaded from the JGI *Populus trichocarpa* database (Supplementary Table 2). Domain detection showed that plant TLP proteins can be classified into three major classes according their domain organization. Most proteins (30 of 36) contain both a highly conserved F-box domain at the N terminus and a TUB domain at the C terminus. Five proteins (AtSPL8 in *Arabidopsis*, OsTLP4 in rice, and PtTLP1, PtTLP7, and PtTLP10 in poplar) contain only the TUB domain at the C terminus. AtTLP4 was the only protein that contained two short TUB domains in the middle.

Arabidopsis TLP genes were dispersed on all the chromosomes except chromosome 4. Seven *TLP* genes were found on chromosome 1, two on chromosome 2, and one each on chromosomes 3 and 5, respectively. The 14 rice *TLP* genes were present on 9 of 12 rice chromosomes. One rice *TLP* gene each was located on chromosomes 3, 4, 7, 8, 11, and 12, two on chromosome 2, and three each on chromosomes 1 and 5, respectively. There are 19 chromosomes in the poplar genome. In our analysis, two poplar *TLP* genes were localized to a scaffold that had not been mapped on chromosomes. Among the other poplar genes that were mapped, one gene each was located on chromosomes 1, 7, 8, 10, and 11, and two each on chromosomes 2 and 5.

Phylogenetic relationships of TLP family in Arabidopsis, rice, and poplar

To investigate the molecular evolution and phylogenetic relationships among *TLPs* in *Arabidopsis*, rice, and poplar, three combined phylogenetic trees were constructed with the neighbor-joining (NJ), minimum evolution (ME), and maximum parsimony (MP) methods, respectively. They exhibited the same topology. We selected only the NJ tree for further analysis, as it was supported by the highest bootstrap values. The NJ phylogenetic tree divided the plant *TLP* genes into three distinct subfamilies: A, B, and C (Fig. 1). Subfamily B contained only one member, AtTLP4 in *Arabidopsis*; its domain structure was far different from those of other members of this family, suggesting that *AtTLP4* independently evolved in the *Arabidopsis* genome. The alternative explanation is that members of this subfamily in rice and poplar were lost during the long evolutionary period. Subfamily C contained three proteins: AtTLP8 in *Arabidopsis*, OsTLP4 in rice, and PtTLP1 in poplar. One apparent feature of all three proteins was a TUB domain at the C terminus. These three genes, therefore, may have originated from one gene in an ancestral species, and did not expand after the split between dicot and monocot. All proteins in subfamily A contained both highly conserved TUB and F-box domains except for PtTLP7 and PtTLP10. The proteins in subfamily A were further divided into four distinct orthologous clusters (OCs): A1–A4. These four OCs all contained *Arabidopsis*, rice, and poplar TLP proteins, indicating that the main characteristics of this subfamily in plants were established before the dicot-monocot split.

Fourteen pairs of paralogous genes (three for *Arabidopsis*, six for rice, and five for poplar) were identified at the terminus of the phylogenetic tree. All paralogous genes belonged to subfamily A. This result indicates that most of the *TLP* genes belonging to subfamily A in *Arabidopsis*, rice, and poplar had expanded in a species-specific manner, and probably only a few members originated from the common ancestral genes that existed before the divergence of monocot and dicot. The species differed with respect to expansion of the four OCs; for example, the genes in rice and poplar had expanded in OC A4, but the *Arabidopsis* gene in OC A4 had not expanded. It is also interesting that all of the poplar *TLP* genes in subfamily A were followed in paralogous pairs.

Sequence alignment of the TLP proteins in Arabidopsis, rice, and poplar

All subfamily A proteins contained conserved TUB domains, and most of them also contained highly conserved F-box domains except for PtTLP7 and PtTLP10. Genes *PtTLP7* and *PtTLP10* were demonstrated to be paralogous to each other, and orthologous to *AtTLP7* in *Arabidopsis*, whereas the latter possessed an F-box domain in the middle of its sequence. So it may be deduced that the F-box domain was lost during the long evolutionary period of proteins PtTLP7 and PtTLP10. We aligned all the amino acid sequences of proteins in subfamily A (Supplementary Fig. 1), and found that there were four conserved blocks for the TUB domain in all the protein sequences. PROSITESCAN (http://www.expasy.ch/tools/scanprosite/) was used to search the PROSITE database for functional motifs. Most proteins in subfamily A contained two signature patterns called the TUB1 and TUB2 motifs (Fig. 1). These two motifs were all located at the C termini of TUB domains and contained 14 and 16 amino acid residues, respectively. Although the similarity of the TUB domains was not very high, we found some highly conserved blocks in these domains. For instance, a motif in the middle of the domains with amino acid residues PGPRRM was highly conserved in all proteins, which could be another feature of TUB domains.

The sequences connecting the F-box and TUB domains were also found to be conserved. The sequence length for this segment in all two-domain-containing proteins of this subfamily was 10 amino acid residues, except in OsTLP10, where it was 56 residues, suggesting that the F-box and TUB domains and the sequence connecting them should be evolutionarily conserved among plants. No significant conserved sequences were detected at the N termini of proteins in this family.

Analysis of the intron distributions of the TLP genes in Arabidopsis, rice, and poplar

The intron distribution can also provide important evidence to support phylogenetic relationships in a gene family [15]. In the TLP family, genes in subfamily C were shown to possess more introns than genes in subfamilies A and B (Fig. 1). There were seven introns in the rice gene *OsTLP4*, whereas eight introns were found in *Arabidopsis* gene *AtTLP8* and poplar gene *PtTLP1*. Comparison of the sequences revealed that a redundant intron located at the N terminus had truncated the first exon into two small exons in *Arabidopsis* and poplar. This intron was found only in dicots and might have been gained after the split between dicot and monocot. The majority of the genes in OCs A1, A2, and A4 had three introns (22/25) and their positions were highly conserved. All genes in OC A3 contained four introns and their positions were also conserved. The main difference between the genes in A3 and those in the other OCs was that a redundant intron located at the C terminus had truncated the last