Plant Syst Evol
DOI 10.1007/s00606-009-0221-5

ORIGINAL ARTICLE

Molecular phylogeny of miR390-guided *trans*-acting siRNA genes (*TAS3*) in the grass family

Dan Shen · Sheng Wang · Huan Chen · Qian-Hao Zhu · Chris Helliwell · Longjiang Fan

Received: 13 January 2009 / Accepted: 3 September 2009

Abstract *Trans*-acting siRNAs (tasiRNAs) are a plant-specific class of 21-nt endogenous siRNAs that function as miRNA-like posttranscriptional negative regulators. Several tasiRNA loci (known as *TAS* genes) have been characterized to date in rice and *Arabidopsis*. The *TAS3* family is distinguished from other *TAS* loci by the dual miR390 complementary sites flanking the tasiRNA region and its dependence on ARGONAUTE7. In this study, 55 putative *TAS3* genes were identified by database mining and PCR amplification from the grass family. Phylogenetic analysis indicated that several genome/gene duplication events have been involved in the expansion of *TAS3* genes while some *TAS3* genes might have been lost during evolution of the grass family. The role of miR390 target genes in the origin of *TAS3* genes are discussed.

Keywords *Trans*-acting siRNA (tasiRNA) · *TAS3* · Poaceae · Evolution · miR390

Electronic supplementary material The online version of this article (doi:10.1007/s00606-009-0221-5) contains supplementary material, which is available to authorized users.

D. Shen · S. Wang · H. Chen · L. Fan (✉)
Institute of Crop Science and Institute of Bioinformatics, Zhejiang University, 310029 Hangzhou, China
e-mail: fanlj@zju.edu.cn

S. Wang
Hangzhou Dianzi University, Xiasha, 310018 Hangzhou, China

Q.-H. Zhu · C. Helliwell
CSIRO Plant Industry, Canberra, ACT 2601, Australia

Introduction

Most eukaryotes contain RNA-based silencing pathways that direct transcriptional and posttranscriptional activities to negatively regulate genes, repetitive/mobile elements, and viruses (Kasschau et al. 2007). Small RNAs play a major role in these pathways. The small RNA population of plants is composed of 20- to 24-nucleotide (nt) microRNAs (miRNAs) and several classes of small interfering RNAs (siRNA), each derived from distinct modes of biogenesis and genomic loci (Rajagopalan et al. 2006). miRNAs and siRNAs are distinguished by the structure of their precursors and by their targets. miRNAs are derived from the short, imperfectly paired stem of a much longer foldback transcript by the activity of Dicer or Dicer-like (DCL) ribonucleases and regulate the expression of transcripts with which they may have limited similarity. siRNAs arise from a long, perfectly paired double-stranded RNA (dsRNA) generated by RNA-dependent RNA polymerase (RDR) or other mechanisms and typically direct the cleavage of transcripts to which they are completely complementary, including the transcript from which they are derived (Yoshikawa et al. 2005). All forms of silencing by small RNAs require an effector protein of the Argonaute (AGO) family.

Trans-acting siRNAs (tasiRNAs) are a class of 21-nt endogenous siRNAs found in plants. tasiRNAs function in a similar manner to miRNAs, cleaving target mRNAs via interaction with a target site with nonperfect complementarity. Each tasiRNA locus (known as a *TAS* gene) produces a noncoding transcript, which is first cleaved by miRNAs. The 5′ or 3′ product of the cleaved transcript is then used to generate mostly 21-nt tasiRNAs through the SGS3/RDR6/DCL4 pathway, in which SGS3 stabilizes the cleaved product that is converted to dsRNA by RDR6. The resulting

Published online: 22 September 2009

dsRNA is processed into 21-nt siRNA duplex in register with the miRNA cleavage site. One strand of the siRNA duplex is selectively sorted to one of the AGO proteins to guide cleavage of target mRNAs (Allen et al. 2005; Yoshikawa et al. 2005). miRNA-guided cleavage of *TAS* transcript is important because siRNA produced in most other registers from a *TAS* locus would not have sufficient homology to direct cleavage of target mRNAs (Axtell et al. 2006). *TAS* genes were first described in *Arabidopsis* (Peragine et al. 2004; Vazquez et al. 2004) where to date 10 loci from four *TAS* families (*TAS1–TAS4*) have been shown to generate tasiRNAs (Peragine et al. 2004; Vazquez et al. 2004; Allen et al. 2005; Rajagopalan et al. 2006; Howell et al. 2007). The phasing register of *TAS1* and *TAS2* (targeted by miR173) and *TAS4* (targeted by miR828) is determined by a single miRNA binding site, whereas *TAS3* requires dual miR390 binding sites to define the region for tasiRNAs production (Allen et al. 2005; Yoshikawa et al. 2005; Rajagopalan et al. 2006; Montgomery et al. 2008).

The *TAS3* family is distinct from other identified *TAS* loci in at least three ways: (1) Dual miR390 binding sites are functionally required in dicot (*Arabidopsis*) and monocot (rice) plants, and even in gymnosperms (*Pinus*) and the moss *Physcomitrella patens*. In addition to an miR390 binding site at the 3′ end, *TAS3* requires another miR390 binding site near the 5′ end of the *TAS* transcript. (2) tasiRNAs from the *TAS3* locus, but not those from *TAS1* or *TAS2*, are dependent on AGO7 that is specifically associated with miR390 (Montgomery et al. 2008). (3) The cleavage guide, miR390, is conserved across the plant kingdom (Embryophyta).

Biological functions have been assigned to a *TAS3* family tasiRNA, tasiARF. tasiARF downregulates mRNAs encoding *auxin response factors* (*ARFs*), including *ARF2*, *ARF3*, and *ARF4* (Allen et al. 2005; Williams et al. 2005), and is involved in the proper timing of vegetative shoot development and establishment of leaf polarity (Hunter et al. 2006). In *Arabidopsis*, *TAS3a*, with two near-identical 21-nt tasiARFs that coaligned with the phases D7(+) and D8(+), was first identified by Allen et al. (2005); two additional *TAS3* family members with a single tasiARF at D8(+) or D4(+) positions were subsequently found (Howell et al. 2007). *TAS3* loci have been reported in rice and other seed plants based on their conserved tasiARFs and dual miR390 complementary sites (Allen et al. 2005; Williams et al. 2005; Axtell et al. 2006; Zhu et al. 2008). In addition, at least four *TAS3* genes have been experimentally identified in mosses (Talmor-Neiman et al. 2006; Axtell et al. 2006). Computational procedures have also been successfully used to identify miRNAs based on evolutionary conservation (Jones-Rhoades and Bartel 2004). The known *TAS3* genes all contain single or tandem putative tasiARFs in 21-nt phasing registers, although tasiARF sequences are not conserved between mosses and seed plants (Axtell et al. 2007).

Several *TAS3* genes have been identified in rice (Allen et al. 2005; Williams et al. 2005; Axtell et al. 2006; Liu et al. 2007; Zhu et al. 2008). High conservation of *TAS3* between rice and other plants was observed, and this provides an opportunity to identify putative *TAS3* genes through sequence similarity search and PCR amplification. In this study, putative *TAS3* genes in the grass family were identified through both data mining in current databases and PCR amplification, and their phylogeny and evolution are then discussed.

Materials and methods

Plant materials

Leaves of maize (*Zea mays*, var. B73), barley (*Hordeum vulgare*, var. 03-34), wheat (*Triticum aestivum*, var. Taber), sorghum (*Sorghum bicolor*, var. 121B), and sugar cane (*Saccharum officinarum*, var. Zhihongpi), which were provided by the Institute of Crop Science and Institute of Nuclear-Agricultural Science, Zhejiang University, were used for DNA extraction.

Computational identification of *TAS3* genes

To gain a more complete understanding of the *TAS3* family and its evolution in plants, an exhaustive BLAST search and model search (HMM-profile) using four rice *TAS3* genes as queries was done against current public nucleotide databases (GenBank) to find putative *TAS3* genes in other grass species. A putative *TAS3* gene must meet the following criteria: (1) it must contain two miR390 complementary sites; (2) it must have at least one conserved 21-nt tasiARF; (3) the space between the two miR390 complementary sites must be less than 273 nt (13 phases) (a typical known *TAS3* gene has a space size of 231 nt or 11 phases between two miR390 complementary sites); (4) because of potential sequencing errors in single pass sequences, if two or more putative *TAS3* candidates were found, they were treated as unique *TAS3* genes only when they had at least two nucleotide differences in the sequences between two miR390 complementary sites unless located on different chromosomes.

PCR amplification and sequencing

Two pairs of primers were designed for the one- and two-tasiARF *TAS3* genes, based on the conserved tasiARF and the 3′ miR390 complementary site of the *TAS3* gene identified by the database mining (Table 1). The primer pairs

AR1F/AR1R (5′TTCTTGACCTTGCAAGACT3′/5′AGCTCAAGAGGGATAGACG3′) and AR2F/AR2R1 or AR2R2 (5′CCTTTTCTTGACCTTGTAAG3′/5AGCTCAGAAGGGATAGA3′ or 5′AGCTCAGGAGGGATAGAAAGG3′) were used to amplify the one- and two-tasiARF *TAS3* genes in maize, barley, wheat, sorghum, and sugar cane. The expected length of the PCR products was about 90–160 nt. Purified PCR products were cloned into pGEM-T vector and at least 10 (10–15) clones were sequenced for each PCR product amplified. In total, 119 sequences were generated from five species and 27 unique putative *TAS3* sequences were deposited into GenBank, with accession numbers EU327115-EU327141.

Sequence analysis

TAS3 sequences identified by database mining were aligned using MAFFT (version 5.8) (Katoh et al. 2005). Neighbor-joining phylogenies based on the Kimura two-parameter distance matrix were generated by MEGA version 3.1 (Kumar et al. 2004). Bootstrap confidence values were obtained from 1,000 replicates. HMMER (Eddy 1998) was used to generate the consensus sequences of HMM profiles of *TAS3* genes.

Results

Identification of *TAS3* genes in the grass family

Using four previously identified rice *TAS3* genes (Zhu et al. 2008) as query sequences, at least 36 putative *TAS3* genes were found from other members of the grass family by sequence similarity (BLASTN) and HMM-based model search in current databases (Table 1). As in rice, *TAS3* genes in other grass families can also be classified into two

Table 1 Putative *TAS3* genes identified in the grass family (Poaceae)

Species	Abbreviation	Two-tasiARF *TAS3* (*TAS3a*)[a]		One-tasiARF *TAS3* (*TAS3b*)[a]	
		TAS3a1	*TAS3a2*	*TAS3b1*	*TAS3b2*
O. sativa ssp. *japonica*	Os	OsTAS3a1 (EU293143)	OsTAS3a2 (Os05g04200)	OsTAS3b1 (EU293144)	OsTAS3b2 (EU293145)
O. sativa ssp. *indica*	Os	AAAA02008565	AAAA02015313	AAAA02006907	AAAA02014293
O. nivara	On	CL736635			
O. rufipogon	Or	CL856635	EI099869	CL809151	
O. glaberrima	Og	CW654135			
O. minuta	Om	CZ596929			
O. officinalis	Oo			DU585057	
O. alta	Oa			CZ438264	
O. granulata	Og			DU658183	
Z. mays	Zm	DN227887	BE519095[b, c]	CO527402[c]	
		DV526105		BT017353	
				AC209970	
				AC211697	
H. vulgare	Hv		BF264964[b]		BU970179
T. aestivum	Ta	CN010916[b]	CS592774[c]		CJ606719[c]
		BE442731			
S. bicolor	Sb	CD464142[b, c]		CW402056[c]	
				ER669122[c]	
S. officinarum	So	CA145655[b]	CA150646		
		CA148474[c]			
		CA150593			
		CA150459			
S. australiense	Sa	CS592782			
P. glaucum	Pg			CD724582	
Total number		16	7	13	4

[a] Classification based on rice *TAS3* genes and the phylogenetic trees shown in Fig. 2. *TAS3a* and *b* have two and one conserved 21-nt tasiARFs, respectively

[b] Axtell et al. (2006)

[c] Also identified by PCR amplification in this study

subfamilies, one- and two-tasiARF *TAS3* subfamilies that contain one and two of the tasiARFs, respectively. Of the 36 putative *TAS3* genes, 15 are in the one-tasiARF subfamily and 21 are in the two-tasiARF *TAS3* subfamily. Other than 13 putative *TAS3* loci identified in *indica* and wild rice, most of the 23 *TAS3* loci identified in other grass species were found from EST sequences. Copy numbers of *TAS3* identified differ among species, from two in barley to seven in maize. In maize, at least two copies were found on the same BAC sequence (AC185460), and their EST sequences (CO527402 and BT017353) are available in current databases (Table 1). Moreover, two identical copies were found at two different chromosomes (AC209970 and AC211697 on chromosomes 4 and 5, respectively). Several observations from small RNA deep sequencing and phase conservation support our above annotation.

Large-scale small RNA sequencing has been carried out in several cereals, which provides us an opportunity to confirm the newly identified *TAS3* genes. Small RNA reads generated from the corresponding phases of *TAS3* loci should be present in these datasets where they are of sufficient depth. In maize 36,563 distinct small RNA reads are currently available (http://sundarlab.ucdavis.edu/smrnas/). Eight 21-nt small RNA reads were perfectly ($\leq$1 mismatch) mapped to the phase positions of three annotated *TAS3* genes (BT017353, BE519095, and DN227887) (Table S1), suggesting that these *TAS3* genes are actively generating tasiRNAs.

Highly conserved sequences were observed in the two miR390 complementary sites and in the tasiARF phase(s) (data not shown). In general, higher conservation was observed at the tasiARF and the 3′ miR390 complementary sites than at the 5′ miR390 complementary site, and higher conservation in the one-tasiARF *TAS3* genes than in the two-tasiARF *TAS3* genes. In particular, only one and two substitution sites were observed in the tasiARF phase and in the 3′ miR390 complementary site in the one-tasiARF *TAS3* subfamily. In the two-tasiARF *TAS3* subfamily, high polymorphism was observed in the two ends of the 3′ miR390 complementary sites and at the middle of the 5′ miR390 complementary site, which is similar to the pattern observed in the seed plants by Axtell et al. (2006).

In the case of *Arabidopsis TAS3* gene, it has been shown that the 3′ miR390 target site sets the phase for the production of the tasiARF (Allen et al. 2005; Yoshikawa et al. 2005). The structures of *TAS3* genes are characterized by one or two tasiARF phases flanked by an miR390 complementary site at each side. Each *TAS3* gene can therefore be divided into two parts relative to the tasiARF phase: 5′ and 3′ regions. Sequence lengths of the two regions were investigated and exact differences of 21-nt at the 3′ region were observed between sequences (Fig. S1). In the one-tasiARF *TAS3* subfamily, nucleotide length changes strictly by phase (21-nt) in the 3′ region, while the 5′ region usually has a shift of several nucleotides. All *TAS3* genes identified have a length of 63, 84, or 105-bp, or three, four, and five phases in the 3′ region, i.e., between the tasiARF and the 3′ miR390 binding site. For example, *TAS3b* genes with four and five phases were observed in rice, maize, millet, and sugar cane, while *TAS3* genes with three phases were detected in wheat and barley (Fig. S1). In the two-tasiARF *TAS3* subfamily, a similar pattern was also observed (data not shown). These results indicate a high conservation of phase structure in the region between the tasiARF and the 3′ miR390 complementary site. Our result also implies that the number of phases in *TAS3* genes might differ, which is consistent with the observation of 9, 10, or 11 phases in the known *TAS3* genes in plants (Howell et al. 2007; Zhu et al. 2008).

To find additional new *TAS3* genes and estimate their copy numbers, PCR primers based on the conserved tasiARF and the 3′ end of the miR390 complementary site were used to amplify *TAS3* fragments from maize, wheat, barley, sorghum, and sugar cane. PCR products were excised, purified, and sequenced. In total, 27 unique sequences were identified (Table 2). The length of all 27 sequences was in exact 21-nt phases. Of the 27 *TAS3* fragments, 15 had two tasiARFs and the other 12 had one tasiARF. Eight of these 27 sequences matched *TAS3* genes identified by database mining from five grass species (Table 1). The remaining 19 were new putative *TAS3* sequences. For example, no putative one-tasiARF *TAS3* gene could be found in available sugar cane sequence whereas two were found by PCR amplification. The inconsistency between computational prediction and PCR amplification might be because the genome sequencing of most species used in this study is not complete, resulting in some *TAS3* genes that were not predicted but amplified. Altogether, database mining and PCR amplification identified 56 putative *TAS3* genes in the grasses, with 5–11 putative *TAS3* genes in each species. Each species has multiple copies of one- and two-tasiARF *TAS3* genes.

Genomic synteny

The Sorghum Genome Project (Paterson et al. 2009; ftp://ftp.jgi-psf.org/pub/JGI_data/Sorghum_bicolor/v0/, Sbi0 Assembly) provides an opportunity to investigate genomic synteny of *TAS3* genes between members of the grass family. High genomic synteny was observed in the segments containing *TAS3* genes in rice and sorghum, for example, *TAS3b1* (Fig. 1a) and *TAS3a1* (Fig. S2A).

Rice underwent a genome duplication about 70 million years ago, and many syntenic regions between chromosomes from the duplication event are retained in the rice genome (Paterson et al. 2004; Guyot and Keller 2004;

Table 2 Estimation of copy numbers of *TAS3* genes in the grass family through database mining (*1*) and PCR amplification (*2*)

Sub-family	Method	*H. vulgare*	*T. aestivum*	*Z. mays*	*S. bicolor*	*S. officinarum*	*O. sativa*[a]
Two-tasiARF	1	1	3	3	1	5	–
	2	1	4	4	3	3	–
	Total (unique)	2	6	6	3	7	2
One-tasiARF	1	1	1	4	2	0	–
	2	2	4	2	2	2	–
	Total (unique)	3	4	5	2	2	2
Total (unique)		5	10	11	5	9	4

[a] Based on Williams et al. 2005 and Zhu et al. 2008

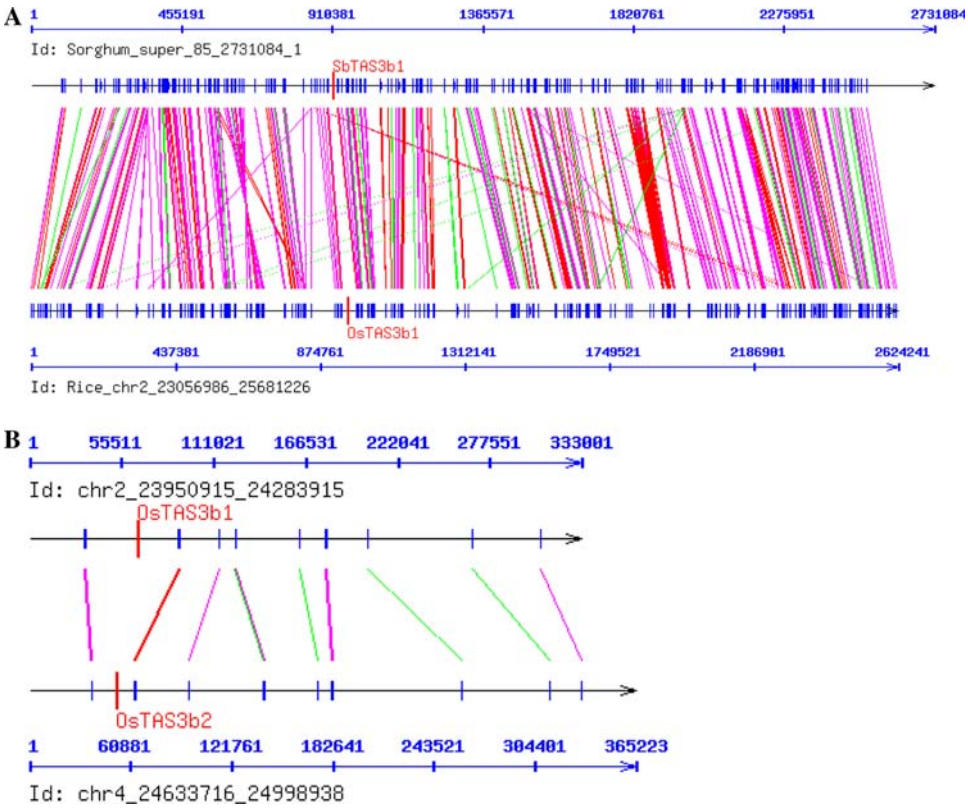

Fig. 1 Genomic synteny of *TAS3* genes. **a** Genomic synteny of *TAS3b1* genes between rice and sorghum. A 2.8-Mb sorghum genomic segment containing *TAS3b1* and the syntenic sequences in rice are shown. **b** Alignments of duplicated segmental pairs harboring the *TAS3b* subfamily members in rice

Zhang et al. 2005; Wang et al. 2005; Yu et al. 2005). *OsTAS3b1* and *b2* are located on chromosomes 2 and 4, respectively, and they fall in a syntenic region from the duplication (Fig. 1b), suggesting that they originated from the genome duplication. *OsTAS3a1* and *a2* seem to have been separated before the genome duplication (see below). However, no paralogs of *OsTAS3a1* and *a2* could be detected at their syntenic region in the current rice genome. For example, *OsTAS3a1* is located at chromosome 3, but no homologous sequence can be found in the corresponding 124-kb region on chromosome 7 (29212839-29337147; Fig. S2B). This result suggests that their paralogs might have been lost during evolution.

Phylogenetics of *TAS3* genes

The *TAS3* genes of the grass family identified by this study (Table 1) were used to construct *TAS3* phylogenetic trees with *Arabidopsis* as the outgroup (Fig. 2). One-tasiARF *TAS3* genes of the grass family clearly belong to two sub-branches where *OsTAS3b1* and *b2* are located, respectively (Fig. 2a). As shown by analysis of genomic synteny, the two sub-branches (named *TAS3b1* and *b2*) arose from the genome duplication event, which occurred before the divergence of the grass family but after the monocot–dicot divergence. The two-tasiARF *TAS3* genes also have two main sub-branches or groups where *OsTAS3a1* and *a2* are

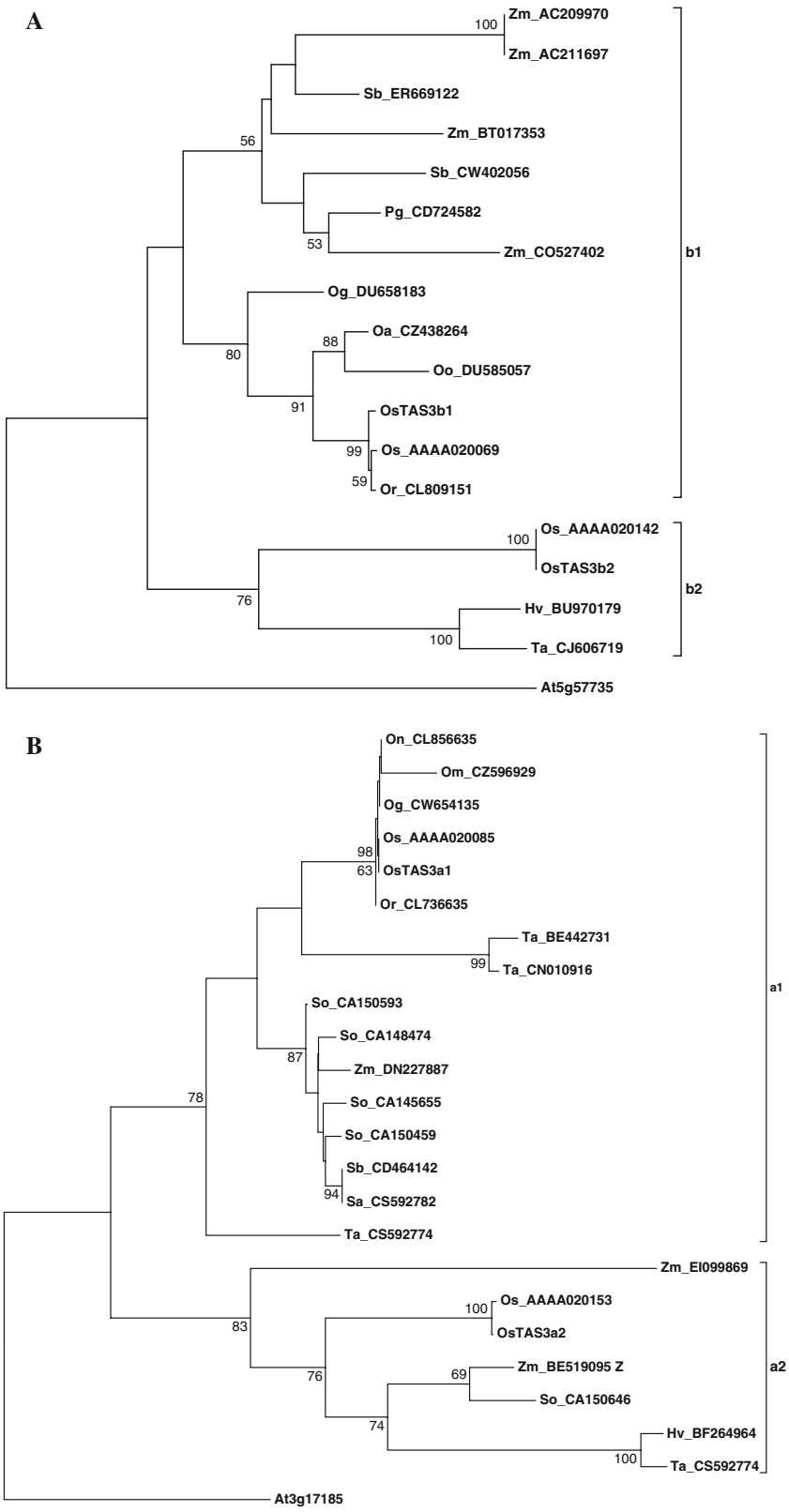

Fig. 2a, b Phylogenetic trees of *TAS3* genes in the grass family. Two *Arabidopsis TAS3* genes are outgroups for the one- (**a**) and two-tasiARF (**b**) *TAS3*. The neighbor-joining trees were generated by MEGA (Kumar et al. 2004) with 1,000 bootstraps, and the bootstrap values are indicated *at nodes* with at least 50% support. For abbreviations of species, see Table 1

located, respectively (Fig. 2b). Beyond rice, putative *TAS3a* genes could be identified in each sub-branch *(TAS3a1* and *a2*), suggesting that the duplication of *TAS3a1* and *a2* should have occurred before the divergence of the grass family. Our other investigation using *Pinus* as the outgroup suggested that these two sub-branches might

have diverged before the dicot-monocot divergence (data not shown). In the *TAS3a1* and *TAS3b1* sub-branches, maize and some other species have more than two *TAS3* copies. The phylogenetic trees suggest that they might have arisen from other recent duplication events. Several studies have suggested that maize and other members of Pooideae have experienced a more recent genome duplication after the genome duplication ~70 million years ago (e.g., Gaut and Doebley 1997). Taken together, our results suggest several genome and/or gene duplication events involved in the expansion of the *TAS3* gene in the grass family.

Discussion

In this study, we identified 55 putative *TAS3* genes from the grass family through database mining and PCR amplification. We found 5–11 copies of *TAS3* genes in each species, including at least two one- and two-tasiARF *TAS3* genes. *TAS3* genes containing one or two tasiARFs have been observed in *Pinus*, suggesting that both of them are ancient. Duplications play an important role in the expansion of *TAS3* genes in the grass family. Several genome and/or gene duplication events were found to be involved in the expansion of *TAS3* genes in the grass family. However, some species have only two *TAS3* genes according to our computational identification and PCR amplification, suggesting that some *TAS3* genes might have been lost during their evolutionary processes. For example, no additional copy of *TAS3a1* or *TAS3b1* was detected in barley according to our database mining and PCR amplification. Compared with dicots, it seems that the grass family contains more copies of *TAS3* genes. *Brassicaeae* (including *Arabidopsis thaliana*, *Brassica rapa*, and *Medicago truncatula*) contains only one copy of the one- and two-tasiARF *TAS3* genes, while *Solanum lycopersicum*, *Populus trichocarpa*, and *Nicotiana tobacum,* for which all or part of the genome sequences is available, harbor only the two-tasiARF *TAS3* gene, suggesting that one-tasiARF *TAS3* subfamily might have been lost in these dicots. Our results demonstrate an evolutionary process of frequent birth and death of *TAS3* genes in the grass family and other plants, similar to observations for *MIRNA* genes in *Arabidopsis* (Fahlgren et al. 2007) and rice (Guo et al. 2008).

The origin of *TAS* genes is not clear compared to *MIRNA* genes, where there is compelling evidence of an evolutionary origin from inverted duplicated segments of protein-coding target genes (Allen et al. 2004). In our computational scan to identify putative *TAS3* genes, we found some loci that do not satisfy our criteria for *TAS3* genes but have structural similarity to *TAS3* genes. In particular, we found a rice LRR (leucine-rich repeat) kinase gene (Os02g10100) that has dual miR390-targeted sites (Fig. S3), and its miR390-guided cleavage of the 3′ miR390 complementary site has been validated (Sunkar et al. 2005). No small RNA reads in the current small RNA databases could be mapped to this locus, suggesting that this locus may not be active. Similar loci can also be found in other cereals, such as maize (CC731078) and sorghum (Super_11, http://www.phytozome.net/sorghum). Based on the targeting of the PPR-P clade by miRNA and RDR6/DCL4 systems and the resulting pattern of phased tasiRNAs produced, Howell et al. (2007) speculated that *TAS* loci may originate from miRNA-targeted protein-coding genes. PPR-P genes belong to a rapidly expanding gene clade from which many 21-nt phased small RNAs are generated and, in turn, these 21-nt small RNAs are post-transcriptional regulators of the PPR-P genes. These PPR-P loci look like known *TAS* genes, except that most of them still encode functional proteins. A nascent *TAS* locus is believed to be retained if suppression of gene-of-origin family members provided an advantage (Howell et al. 2007). Just like PPR-related genes, there are hundreds of LRR domain-containing genes (such as LRR kinase, NBS-LRR) in the rice genome (Goff et al. 2002). Some of them, such as *Os02g10100*, might evolve into a functional *TAS* locus after acquisition of an miRNA cleavage site if selection pressure favors the generated phased siRNAs. Whether or not phased siRNAs are generated from the *TAS3*-like LRR-containing locus needs further investigation using small RNA deep sequencing.

Acknowledgments This work was supported by the National High Technology Research and Development Program of China (2006AA10A102), National Basic Research Program of China (2006CB101700), and the Research Foundation for Doctoral Programs in China to L.F. Thanks the PPP project by China Scholarship Council ([2007]3086) and DAAD (D/08/01754).

References

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nature Genet 36:1282–1290

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121:207–221

Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell 127:565–577

Axtell MJ, Snyder JA, Bartel DP (2007) Common functions for diverse small RNAs of land plants. Plant Cell 19:1750–1769

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL et al (2007) High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of miRNA genes. PLoS ONE 2:e219

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 296:92–100

Guo X, Gui Y, Wang Y, Zhu Q-H, Helliwell C, Fan L (2008) Selection and mutation on microRNA target sequences during rice evolution. BMC Genomics 9:454

Guyot R, Keller B (2004) Ancestral genome duplication in rice. Genome 47:610–614

Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC (2007) Genome-wide analysis of the RNA-dependent RNA polymerase6/DICER-like4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell 19:926–942

Hunter C, Willmann M, Wu G, Yoshikawa M, de la Luz Gutiérrez-Nava M, Poethig S (2006) Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. Development 133:2973–2981

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799

Kasschau KD, Fahlgren N, Chapman E, Sullivan C, Cumbie J, Givan S, Carrington J (2007) Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol 5:e57

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucl Acids Res 33:511–518

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinf 5:150–163

Liu B, Chen Z, Song X, Liu C, Cui X, Zhao X, Fang J, Xu W, Zhang H, Wang X et al (2007) *Oryza sativa* dicer-like4 reveals a key role for small interfering RNA silencing in plant development. Plant Cell 19:2705–2718

Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133:128–141

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Jane Grimwood, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A et al (2009) The sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. Genes Dev 18:2368–2379

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–3425

Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice. Plant Cell 17:1397–1411

Talmor-Neiman M, Stav R, Klipcan L, Buxdorf K, Baulcombe DC, Arazi T (2006) Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J 48:511–521

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P (2004) Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Molecular Cell 16:69–79

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Williams L, Carles CC, Osmont KS, Fletcher JC (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis* ARF2, ARF3, and ARF4 genes. Proc Natl Acad Sci USA 102:9703–9708

Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. Genes Dev 19:2164–2175

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C et al (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zhang Y, Xu G, Guo X, Fan L (2005) Two ancient rounds of polyploidy in rice genome. J Zhejiang Univ Science B 6:87–90

Zhu Q-H, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18:1456–1465