

M. Saito · J. Yonemaru · G. Ishikawa · T. Nakamura

A candidate autonomous version of the wheat MITE *Hikkoshi* is present in the rice genome

Received: 17 November 2004 / Accepted: 9 March 2005 / Published online: 10 May 2005
© Springer-Verlag 2005

Abstract A miniature inverted-repeat transposable element (MITE), designated as *Hikkoshi*, was previously identified in the null *Wx-A1* allele of Turkish bread wheat lines. This MITE is 165 bp in size and has 12-bp terminal inverted repeats (TIRs) flanked by 8-bp target site duplications (TSDs). Southern and PCR analyses demonstrated the presence of multiple copies of *Hikkoshi* in the wheat genome. Database searches indicated that *Hikkoshi* MITEs are also present in barley, rice and maize. A 3.4-kb element that has *Hikkoshi*-like TIRs flanked by 8-bp TSDs has now been identified in the rice genome. This element shows high similarity to the 5' subterminal region of the wheat *Hikkoshi* MITE and contains a transposase (TPase) coding region. The TPase has two conserved domains, ZnF_TTF and hATC, and its amino acid sequence shows a high degree of homology to TPases encoded by *Tip100* transposable elements belonging to the *hAT* superfamily. We designated the 3.4-kb element as *OsHikkoshi*. Several wheat clones deposited in EST databases showed sequence similarity to the TPase ORF of *OsHikkoshi*. The sequence information from the TPase of *OsHikkoshi* will thus be useful in isolating the autonomous element of the *Hikkoshi* system from wheat.

Keywords Miniature inverted-repeat transposable elements (MITEs) · Autonomous element · Transposase · *Triticum aestivum* · *Oryza sativa*

Introduction

Class II transposable elements (or DNA TEs) have terminal inverted repeats (TIRs) at both ends and are able to transpose into other regions of the genome through DNA intermediates, using a “cut-and-paste” system. Autonomous elements encode a transposase (TPase) that catalyzes their transposition, whereas non-autonomous elements retain the TIRs but do not encode a TPase, and therefore require the presence of an autonomous element for transposition. Maize *Dissociation (Ds)* was the first non-autonomous class II TE to be identified. Transposition of *Ds* requires the autonomous element *Activator (Ac)* (reviewed in Kunze and Weil 2002).

When TPases insert TEs into target DNA, target site duplications (TSDs) are formed on both sides of the TIRs. Classification of class II TEs into superfamilies is based on the sequence similarities of TPases and TIRs and on the length of the TSDs produced. The *hAT* superfamily TEs *Ac/Ds* of maize, *hobo* of *Drosophila melanogaster* and *Tam3* of *Antirrhinum majus* have similar TPases and TIRs, and are associated with 8-bp TSDs (Calvi et al. 1991).

Miniature inverted-repeat transposable elements (MITEs) are very short, non-autonomous, class II elements. While copy numbers of other class II elements are relatively low, MITEs are often found in high copy numbers in their host genome. The first reported MITE was identified as a small insertion in the *waxy (Wx)* gene of maize, and was subsequently shown to belong to a large family of MITEs called *Tourist* (Bureau and Wessler 1992). Since then, researchers have identified various types of MITEs in many plant and animal genomes. MITEs are characterized by their very short

Communicated by M.-A. Grandbastian

M. Saito · G. Ishikawa · T. Nakamura (✉)
Department of Crop Breeding,
Tohoku National Agriculture Research Center,
4 Akahira, Shimo-Kuriyagawa, Morioka, Iwate,
020-0198, Japan
E-mail: tnaka@affrc.go.jp
Tel.: +81-19-6433514
Fax: +81-19-6433514

J. Yonemaru
Department of Animal Production and Grasslands Farming,
Tohoku National Agriculture Research Center,
4 Akahira, Shimo-Kuriyagawa, Morioka,
Iwate 020-0198, Japan

lengths (<600 bp) and the possession of TIRs flanked by TSDs. They have been classified into several superfamilies based on the structural features of their TIRs and TSDs (Feschotte et al. 2002).

Several studies have demonstrated the active transposition of the MITE *mPing* in the rice genome (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003), but the mechanisms responsible for the mobilization and amplification of MITEs remain poorly understood. Since MITEs lack coding sequences, their mobilization requires TPase encoded in *trans*, and autonomous elements related to specific MITEs have been identified. For example, the maize *Tourist*-like MITE *miniature PIF* (*mPIF*) shares several features with the autonomous element *PIFa*, including identical TIRs, similar subterminal sequences and similar target site preferences (Zhang et al. 2001).

Common wheat (*Triticum aestivum* L. 2 $n=42$, AABBDD) is hexaploid and carries most genes in triplicate. While a large body of literature exists on class II TEs of maize and rice, studies of wheat TEs are limited. Transposition of TEs into the coding or promoter regions of genes in rice or maize is often reflected in phenotypic changes, whereas the polyploid nature of wheat makes mutational effects hard to detect phenotypically. Recently, genomic sequence analyses in several contiguous gene regions have revealed the existence of class II TEs in the wheat genome (Anderson et al. 2002; Ramakrishna et al. 2002; Wicker et al. 2003a; Kong et al. 2004). At least 3000 copies of TEs of the CACTA superfamily are present in the diploid *T. monococcum* genome, suggesting that TEs may have played an important role in wheat genome evolution (Wicker et al. 2003b). MITEs may also be important in wheat evolution, and may affect the expression of wheat genes. *Stowaway*-like MITEs and several unclassified MITEs have been detected in genic regions of wheat (Bureau and Wessler 1994; Anderson et al. 2001; Wicker et al. 2001; Hessler et al. 2002; Ramakrishna et al. 2002; Li et al. 2004) and *Stowaway*-like and *Waffle* MITEs have been found in matrix-attachment regions that are thought to affect transcription (Anderson et al. 2002).

Recently, draft-genome sequences for rice (Goff et al. 2002; Yu et al. 2002) and finished sequences for several rice chromosomes (Feng et al. 2002; Sasaki et al. 2002; The Rice Chromosome 10 Sequencing Consortium 2003) have been reported, and searches of these databases have led to the identification of an autonomous element associated with the MITE *mPing* (Jiang et al. 2003). Less sequence information is available for wheat, whose genome is 40 times larger than that of rice. Since rice and wheat diverged from a common ancestor, the availability of the rice-genome sequence provides the opportunity to conduct comparative analyses of TEs in wheat and rice.

We previously identified an insertion in Exon 4 of a null *Wx-A1* allele carried by a bread wheat line from Turkey (Saito et al. 2004). The insertion was short (165 bp) with perfect 12-bp TIRs flanked by 8-bp

TSDs, and therefore had the structural characteristics of a MITE. This MITE was designated *Hikkoshi* (Accession No. AY376310). In this study, we characterize and classify *Hikkoshi* elements, and employ bioinformatic analyses to identify a candidate for the autonomous version of this MITE system in rice genome sequences.

Materials and methods

Plant materials

Seed accessions of *T. aestivum* L. were obtained from the gene bank of the National Institute of Agrobiological Sciences (Tsukuba, Japan).

PCR analysis

Genomic DNA was extracted from young leaf tissue using the Nucleon Phytopure Plant DNA Extraction Kit (Amersham Biosciences, Little Chalfont, UK) according to the manufacturer's instructions. The primers 5'-CAGTGGCGTAGCTAGGGG-3' and 5'-GGCGTAGCCAGCCCAATAA-3', were designed on the basis of the TIR and subterminal sequences of the *Hikkoshi* MITE in the *Wx-A1* gene of Turkey 124, and used to obtain *Hikkoshi* elements from the wheat cultivars Chinese Spring and Turkey 124. The PCR products were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA). Inserts were sequenced using a BigDye Terminator v1.1 Cycle Sequencing kit and an ABI Prism 310 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

Sequence analysis

To find *Hikkoshi*-related sequences, database searches were performed using BLAST programs on servers available from the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/BLAST/>, nucleotide sequence databases nr, est, gss, wgs and htgs and peptide sequence database nr) and FASTA programs on servers available from the DNA Data Bank of Japan (DDBJ; <http://www.ddbj.nig.ac.jp/search/fasta-j.html>, nucleotide sequence databases nr, est, gss and htgs) (Pearson and Lipman 1988; Altschul et al. 1997). The FASTA server at the University of Virginia (<http://fasta.bioch.virginia.edu/fasta/>) was used for two sequence comparisons (Pearson and Lipman 1988). Nucleic acid or amino acid sequences were aligned using CLUSTAL W (Thompson et al. 1994) with default parameters. The sequence alignments were displayed using the GeneDoc program (<http://www.psc.edu/biomed/genedoc>; Nicholas et al. 1997).

Phylogenetic analysis

Amino acid sequences containing hAT dimerization domains were selected from the protein domain database Pfam (Bateman et al. 2004). The full-length amino acid sequence of the *Slide* TPase used in the analysis was deduced from its predicted ORFs. Full-length amino acid sequences of TPases were aligned using CLUSTAL W (Thompson et al. 1994). The phylogenetic tree was constructed using the neighbor-joining option in MEGA Version 2.1 (Kumar et al. 2001) with 1000 bootstrap replications.

Results and discussion

Wheat has multiple copies of the *Hikkoshi* MITE

Southern analysis showed that the *Hikkoshi* MITE represents a repetitive sequence in the wheat genome (data not shown). *Hikkoshi* MITEs were found not only in the *Wx-A1* mutants from Turkey, but were also present in wild-type lines. The MITEs *Toya* and *Tripper* have TSDs and TIRs with characteristics similar to those of *Hikkoshi*. Although the overall sequences of these elements are not similar to that of *Hikkoshi*, the TIR sequences of *Toya* both in rice (5'-CAGTGGCGGA-3') and tobacco (5'-CAGAGGCGGA-3'), and of *Tripper* in wheat (5'-CAGTGGCGGAGCTT-3') are very similar, and these MITEs are flanked by 8-bp TSDs, suggesting that these three elements might belong to the same superfamily. The copy number of *Toya* in rice and tobacco genomes is quite low (Nagano et al. 2002; Schenke et al. 2003) while multiple copies of *Tripper* are found in the wheat genome (Li et al. 2004). *Hikkoshi* appears to resemble *Tripper* in this respect.

The MITEs are often structurally related to autonomous elements that encode a TPase. In the *Tourist*-like MITE family *mPIF* of maize, a conserved terminal region is common to an active DNA transposon family called *PIF* and its autonomous element *PIFa* (Zhang et al. 2001). Genome-wide analyses of transposable elements in rice showed that *Mariner*-like TEs and *PIF*/*Pong*-like TEs are responsible for the amplification of *Stowaway* MITEs and *Tourist*-like MITEs, respectively (Feschotte et al. 2003; Zhang et al. 2004). These MITEs have TIRs and subterminal sequences that are similar to those of their controlling autonomous elements.

We therefore attempted to amplify additional *Hikkoshi* sequences, including autonomous elements, from the genomic DNA of the wheat lines Chinese Spring and Turkey 124 using PCR primers derived from the TIR and subterminal sequences of *Hikkoshi*. In both cultivars, products of approximately 200 bp were preferentially amplified. In addition, a minor band of about 400 bp was amplified from Turkey 124. The 200-bp product was composed of several different sequences, each of which carried several nucleotide substitutions relative to the original *Hikkoshi* sequence from the

Wx-A1 gene (Fig. 1a). Some of these fragments were also 10–80 bp shorter than *Hikkoshi*. The 400-bp fragment possessed the same subterminal sequences as *Hikkoshi* but carried additional sequence within the subterminal regions (Fig. 1b). The extra 241 bp in this fragment shows similarity to several wheat EST clones, but these ESTs themselves do not show significant similarity to known proteins. The PCR data confirm the Southern results showing that the *Hikkoshi* element has spread in the wheat genome, and further indicate that modification of the element has occurred during or after its movement. We expected to find some sequences with homology to known TPases, but no such sequence was identified by this method. Zhang et al. (2001) were also unable to isolate a targeted autonomous element from maize using primers derived from the subterminal region of the corresponding non-autonomous element, although the autonomous element was present in the genome. We therefore decided to obtain additional genomic information to aid us in identifying a candidate autonomous element related to the *Hikkoshi* MITE.

Hikkoshi MITEs are present in other species

We performed database searches to detect additional *Hikkoshi* MITEs in wheat and other plant genomes, at first employing the complete sequence of the wheat *Hikkoshi* MITE as a query in both BLASTN and FASTA searches. Twenty-five accessions were identified that included the 12-bp TIRs and 8-bp TSDs typical of the wheat *Hikkoshi* MITE and showed more than 60% similarity to this MITE (Fig. 2). We categorized these sequences as members of the *Hikkoshi* MITE family. These elements were quite homogeneous in size (157–180 bp), which is characteristic of MITEs belonging to the same family.

The *Hikkoshi* MITE family appears to belong to the *hAT* superfamily, which includes a number of other MITEs (Morgan and Middleton 1990; Besansky et al. 1996; Smit and Riggs 1996; Surzycki and Belknap 1999; Lepetit et al. 2000; Tu 2001; Holyoake and Kidwell 2003). TEs belonging to this superfamily have 8-bp TSDs and TIR sequences that are similar to those of *Hikkoshi*.

Four of the 25 accessions carrying *Hikkoshi* MITEs represent EST clones from wheat and barley (Accession nos. CD911985, BQ901405, CD880478 and CA009194), while the remaining accessions represent genomic clones originating from barley (one clone), *japonica* rice (16 clones), wild rice (two clones) and maize (two clones) (Fig. 2). A number of positions in the subterminal regions of the rice, barley and wheat MITEs were well conserved (Fig. 2). Thus, *Hikkoshi*-like elements exist not only in wheat but are distributed throughout the Poaceae.

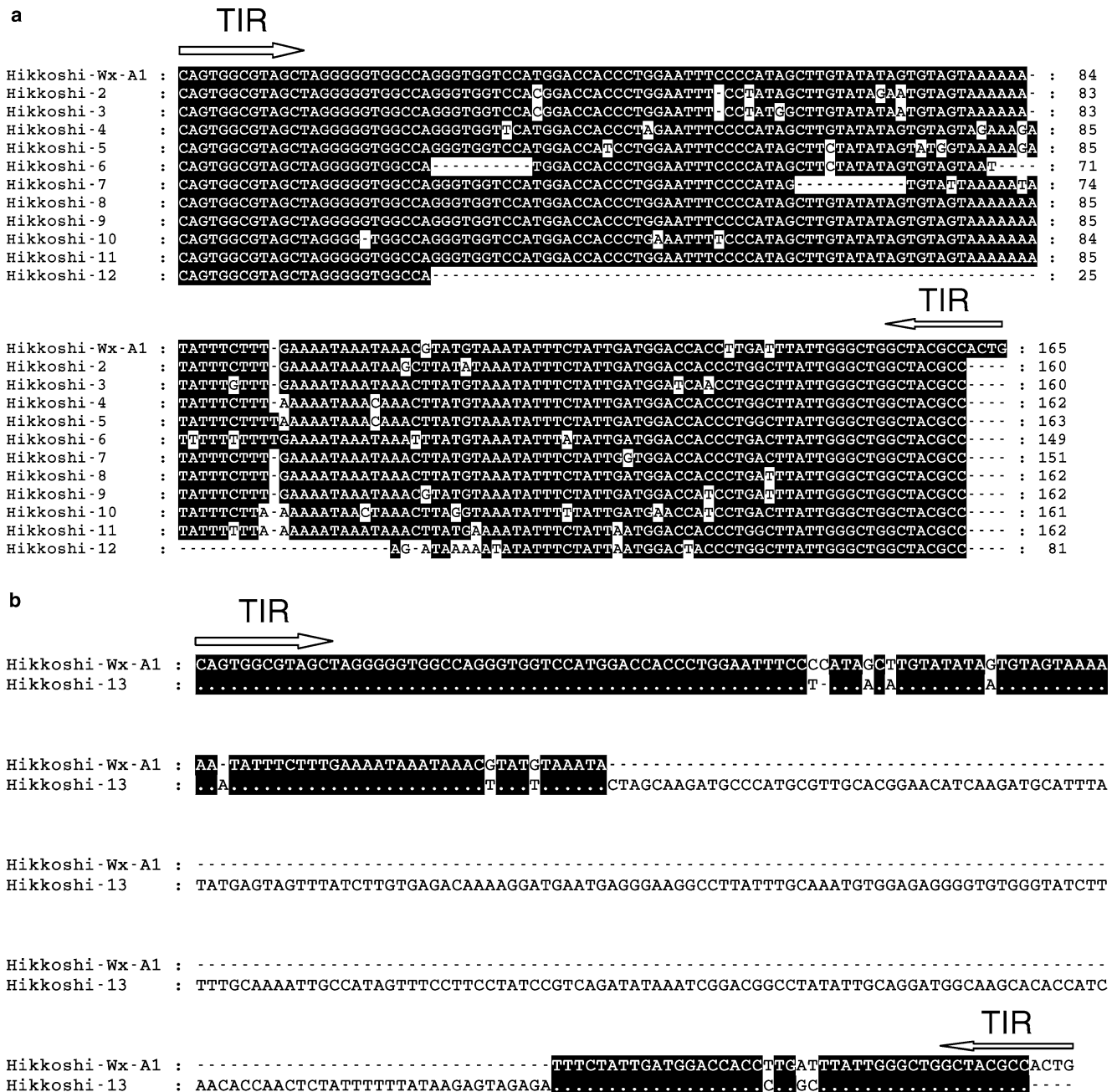
In addition to these 25 elements, five truncated *Hikkoshi* MITEs of 90–158 bp in length and lacking TIRs and subterminal regions at one end were identified in

wheat; one in a cDNA clone (CYP71C3-TA cytochrome P450; AF123603) and four in EST clones. None of the EST clones have been functionally annotated, and their function could not be determined by database searches.

The copy number of *Hikkoshi* MITEs in the genome of *japonica* rice cv. Nipponbare was estimated using BLASTN and FASTA searches. The 16 elements that showed more than 60% similarity to the wheat *Hikkoshi*

MITE (Fig. 2) were employed as queries. Sequences that showed at least 85% similarity over more than 50% of the length of the query were classified as *Hikkoshi* MITEs. By these criteria, a total of 49 elements were identified, including the 16 elements originally identified in *japonica* rice. Among the 33 newly identified elements, 17 elements have TIRs flanked by 8-bp TSDs, three elements lack TSDs and 13 elements did not have a second TIR or subterminal sequences within 250 bp of the first TIR. Five of the 13 elements that appeared to lack TIR and subterminal sequences at one end actually had a second TIR sequence 278–4524 bp downstream from the originally identified TIR. The subterminal sequences flanking the second TIR showed similarity to 3' subterminal sequences of the *Hikkoshi* MITE. Other

Fig. 1 a, b PCR amplification of *Hikkoshi* elements. **a** Multiple alignment of *Hikkoshi* MITEs amplified from the genomic DNA of Turkey 124. **b** Pairwise alignment of the *Hikkoshi* MITE from the *Wx-A1* gene of Turkey 124 and a 400-bp element amplified from the genomic DNA of Turkey 124. The arrows indicate the TIRs. Sequence identities are shaded, and gaps are indicated by dashes



transposable elements, including *Tourist*, *Stowaway*, *Gaijin*, and *En/Spm*, and long terminal repeat (LTR) retrotransposon sequences, were inserted between the two TIRs in several cases (Table 1). *Stowaway*-like and

Tourist-like MITEs are present in high copy numbers in rice; the Nipponbare genome carries an estimated 30,000 copies of *Stowaway*-like and 60,000 copies of *Tourist*-like MITEs (Feschotte et al. 2003; Zhang et al. 2004). Compared with these MITEs, the copy number of the *Hikkoshi* MITE was quite low, and was more similar to the copy numbers of MITEs such as *Toya*, *Akan* and *Mashu* (Nagano et al. 2002). For example, only approximately 10–30 copies of *Toya* are present in the Nipponbare genome (our unpublished data).

Twenty-two of the 49 rice *Hikkoshi* elements were located in the structural regions of putative or hypothetical genes. Eighteen of the remaining 27 elements

Fig. 2 Multiple alignment of *Hikkoshi* MITEs found in databases. The filled arrows indicate TSDs and open arrows indicate TIRs. Conserved nucleotides are shaded. Gaps are indicated by dashes. Accession numbers for the EST or genomic clones carrying these MITEs are shown on the right. The overlines indicate nucleotides in the subterminal regions that are well conserved among rice, barley and wheat. Hv *Hordeum vulgare*, Ta *Triticum aestivum*, Wx *Wx-AL*, Os *Oryza sativa*, Or *Oryza rufipogon*, Zm *Zea mays*

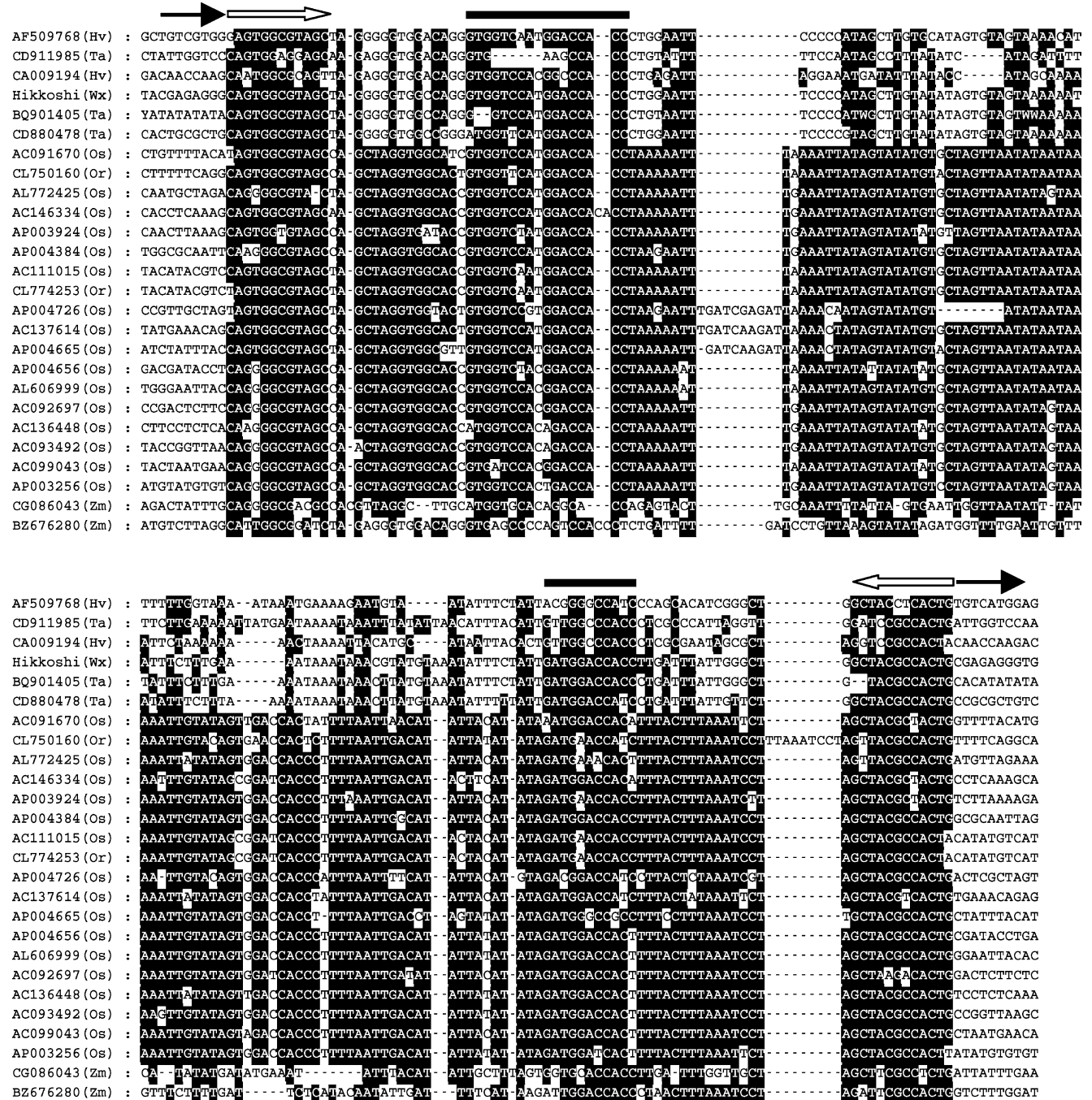


Table 1 *Hikkoshi* MITEs found in *O. sativa* genomic sequences

Accession no.	Position	Size (bp)	Location (bp) ^a	Sequences inserted in MITE
AP003409	135475–135644	170	5' (–1027)	
AP003924	117285–117116	170	In 2	
AP004165	32479–32310	170	In 2	
AC099043	46933–47102	170	In 3	
AC091670	64769–64938	170	5' (–1660)	
AL606999	2098–2267	170	3' (536)	
AC093492	45264–45095	170	3' (1572)	
AC111015	32649–32818	170	3' (469)	
AP004384	11952–11783	170	3' (2160)	
AP004656	47631–47800	170	In 1	
AC092697	40536–40705	170	In 5	
AC136448	66415–66246	170	3' (403)	
AC146334	99694–99523	172	In 1 and Ex 2	
AL844498 ^b	132406–132576	171	5' C	
AL772425	102075–102243	169	3' (5448)	
AP001859	16074–15904	171	In 7	
AP002819	106652–106473	180	3' C	
AP004836	161261–161077	185	5' (–1233)	
AC105731	90014–89832	183	In 4	
AL731642	114722–114902	181	3' (1735)	
AC136524	53089–53268	180	5' (–2689)	
AC137614	14338–14159	180	5' (–1516)	
AC108501	115191–115363	173	In 3	
AP003491 ^b	22536–22717	182	In 2	
AP004726	123130–122960	171	3' C	
AP005292	7656–7835	180	5' (–322)	
AP004665	97236–97413	178	5' (–1814)	
AP005573	8431–8615	185	3' (345)	
AC137595	160271–160449	179	5' (–1140)	
AC148822	51858–52037	180	5' (–2653)	
AC120888	80603–80782	180	5' (–8676)	
AC150702	91253–91078	177	5' (–510)	
AC136377	48656–48744	180	5' C	
AL928774	30989–31168	180	3' (3017)	
AP001389	132928–132781	148	5' (–104)	
AP005769 ^b	89693–89837	145	In 1	
AC084405 ^c	64929–63238	4692	3' C	LTR, simple repeat
AC137699 ^c	9975–9680	296	5' (–692)	<i>Gajjin</i> (127bp)
AC139168 ^c	112724–113001	278	3' (510)	<i>Stowaway</i> (126 bp)
AC136519 ^c	43349–42966	384	In 1	<i>Tourist</i> (100 bp)
AL935154 ^c	56740–58457	1718	5' C	<i>En/Spm</i> (1545 bp)
AP005611 ^d	54053–54219	166	3' (471)	
AC134230 ^d	142445–142593	149	3' C	
AC135907 ^d	91555–91660	106	3' (3456)	
AC091775 ^d	53479–53371	109	3' (2803)	
AC120537 ^d	82696–82829	134	3' C	
AL606687 ^d	140096–139980	117	3' (2350)	
AP003456 ^d	126038–126125	87	In 2	
AP005520 ^d	42480–42362	118	In 1 ^e	

^a Locations of elements: *Ex* exon, *In* intron, 5', 5' flanking region, 3', 3' flanking regions. 5' C or 3' C indicates that the MITE sequence overlaps with the predicted initiation codon or termination codon of the hypothetical protein. Numbers in *parentheses* give the distance of the *Hikkoshi* element from the translation initiation or termination site

^b Elements not flanked by 8-bp TSDs

^c *Hikkoshi* MITEs carrying insertions

^d Defective elements showing similarity with at least one half of the *Hikkoshi* MITE sequence

^e Location classification changes as a result of alternative splicing; may be located either in intron 1 or 2 or in the 5' UTR

were located within 2-kb upstream or downstream of genes (Table 1). This agrees with previous reports suggesting that MITEs are often found in transcribed regions or in regulatory domains of genes (Bureau and Wessler 1994; Bureau et al. 1996; Iwamoto and Higo 2003). It will be interesting to determine whether the presence of the *Hikkoshi* MITEs in rice genes affects the functions of these genes.

OsHikkoshi is the best candidate for the orthologous autonomous element involved in the spread of *Hikkoshi* MITEs

Since we had not obtained a candidate autonomous element from wheat via PCR amplification, we attempted to obtain sequence information for the autonomous element responsible for the mobilization of the

Hikkoshi MITEs of wheat using rice genome bioinformatics. *Hikkoshi* MITEs are present in the rice genome (Fig. 2) and sequence information for the complete rice genome is available, potentially allowing us to identify candidate autonomous elements by database searches. Further detailed surveys were conducted using *Hikkoshi* sequences from both wheat and rice. Good candidates should have complete TPase sequences carrying *Hikkoshi* TIRs and TSDs at both ends.

For these analyses, the wheat *Hikkoshi* sequence was first separated into two sections, the first 80 bp and the last 85 bp, which were used separately as queries for FASTA searches in DDBJ. Although no new accessions having similarity with the 3' half of *Hikkoshi* were

found, 31 new accessions that had high similarity (more than 65%) to the 5' half were identified. In each of these accessions, the region of similarity began with a 12–15-bp motif that was almost identical to the TIR sequence of the *Hikkoshi* MITE. The accessions were then individually scanned for the presence of a second TIR sequence within 5000 bp of this sequence. In 18 of 31 accessions, a second TIR sequence was detected 250–3400 bp downstream of the originally identified TIR. Direct repeat sequences of 8 bp also flanked both TIR sequences in these accessions. The sequences between the TIRs in these accessions were analyzed, and a putative TPase ORF (Accession no. AAP55096) was found in accession AE017121 (Fig. 3a), from chromosome 10 of Nipponbare (The Rice Chromosome 10 Sequencing Consortium 2003). The start site of the putative TPase ORF of *OsHikkoshi* is 753 bp downstream of the 5' TIR (Fig. 3b). This TSD-TIR-TPase-TIR-TSD sequence displays homology not only with the TIRs of the wheat *Hikkoshi* MITE, but also shows approximately 73.3% identity to the 5' subterminal 45-bp region (Fig. 3a). This element was designated *OsHikkoshi* (for *O. sativa Hikkoshi*). Internal deletion derivatives of *OsHikkoshi* were also found in accessions AP003381 (966 bp; positions 114406–113441) and AC135568 (1965 bp; positions 19899–21863). The derivative in AC135568 retained part of the coding sequence of the *OsHikkoshi* TPase. When *OsHikkoshi* was used as a query in a BLAST search against the

Fig. 3 a, b Structure of *OsHikkoshi*. **a** Comparison of wheat *Hikkoshi* MITE and *OsHikkoshi* sequences. The short dotted line indicates the position of an internal 3246-bp *OsHikkoshi* sequence not shown in the alignment. Sequence identities are shaded, TIRs are boxed and TSDs are indicated by arrows. **b** Schematic representation of the structures of *OsHikkoshi* and *Tip100*. The *Tip100* element shown here is from *Ipomoea purpurea* (Accession no. AB004906). Conserved regions, including the ZnF_TTF (black box) and hATC (stippled box) domains, are shaded. The open and filled triangles represent TSDs and TIRs, respectively. TSD sequences are shown in lower case letters and TIR sequences are in capital letters. The numbers above the diagrams refer to the nucleotide position relative to the first base of the 5' TIR. TSD sequences on both sides of *Tip100* are derived from an intron of the *CHS-D* gene in *I. purpurea* (Habu et al. 1998)

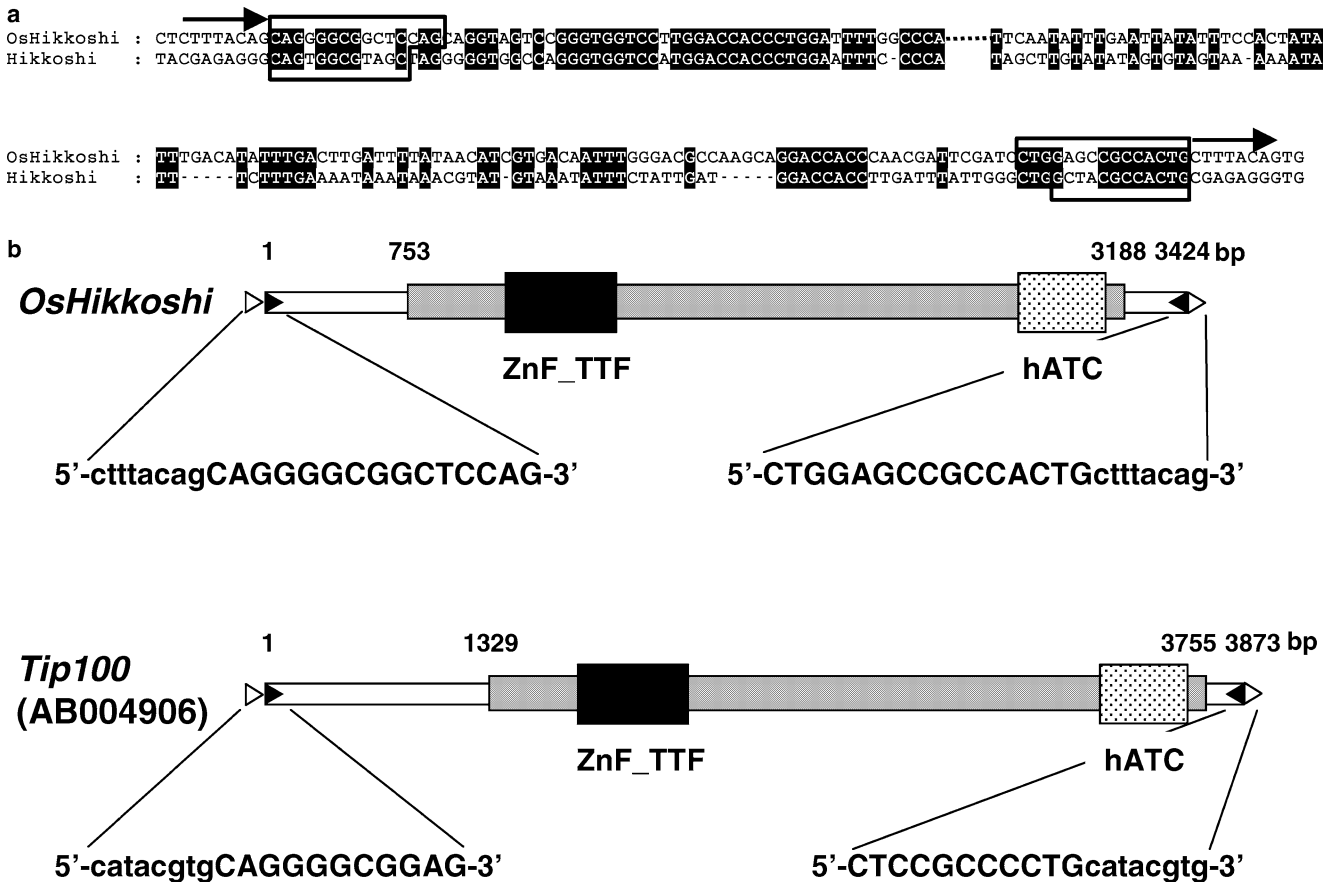


Table 2 Comparison of *Hikkoshi* and *OsHikkoshi* TIRs with TIRs of TPase-encoding elements identified in a BLASTP search

Element or accession no. (source) ^a	Position ^b	5' TIR ^c	3' TIR ^c	Size (bp) ^d
<i>Hikkoshi</i>	<i>Wx-A1</i>	CAGTGGCGTAGC	CAGTGGCGTAGC	165
AAP55096 ^e	AE017121	<u>CAGGGGCGGCTCCAG</u>	<u>CAGTGGCGGCTCCAG</u>	3424
BAC10884 (Os)	AP004361	<u>CAGGGGCGGAGC</u>	<u>CAGTGGCGGAGC</u>	3488
CAE04315 (Os)	AL606658	<u>CAGGGGCGGAGC</u>	<u>CAGTGGCGGAGC</u>	3488
AAP55079 (Os)	AE017121	<u>CAGGGGCGGAGC</u>	<u>CAGTGGCGGAGC</u>	3488
AAF82236 (At)	AC069143	<u>CAGGGACGGAGCCA</u>	<u>CAGTGGCGGAGCCA</u>	3200
AAT38794 (Sd)	AC149267	<u>CAGGGGCGGAT</u>	<u>CAGTGGCGGAT</u>	3872
BAB12027 (Os)	AP002820	<u>CAGTGGCGGAT</u>	<u>CACGGGCGGAT</u>	3796
BAA36225 (Ip)	<i>CHD-D</i> (AB004906)	<u>CAGGGGCGGAG</u>	<u>CAGGGGCGGAG</u>	3873

^a Os *Oryza sativa*, At *Arabidopsis thaliana*, Sd *Solanum demissum*, Ip *Ipomoea purpurea*

^b Genomic clones or genes in which elements are located

^c Completely conserved nucleotides are *underlined*

^d Including TIRs

^e Query sequence for BLASTP search

genomic sequences of Nipponbare, two additional accessions showing similarity with *OsHikkoshi* were identified. The genomic clones AP005925 from chromosome 8 and AC136521 from chromosome 5 appear to have long ORFs, similar to that of *OsHikkoshi*. However, a premature termination codon occurs in the coding region of AP005925 due to a 4-bp insertion, and an insertion of approximately 15,000 bp is found in AC136521 (data not shown). These two elements are thought to be non-autonomous elements derived from *OsHikkoshi*. Thus, *OsHikkoshi* is the strongest candidate for the orthologous autonomous element from which the *Hikkoshi* MITEs were derived, and there is only one intact copy of *OsHikkoshi* in the Nipponbare genome.

However, *OsHikkoshi* does not appear to be a likely candidate for the autonomous element of the *Hikkoshi* MITEs of rice (Table 1) because, contrary to what was observed with the wheat MITEs, the subterminal sequences of the rice MITEs do not show high similarity with subterminal sequences from *OsHikkoshi*. Furthermore, when we searched for autonomous elements using rice *Hikkoshi* MITE sequences, *OsHikkoshi* was not identified as a candidate element in the screen. In fact, no autonomous elements were identified when the rice MITE sequences were used as queries. Thus, MITEs with high similarity to *OsHikkoshi* are found in the wheat genome but not in the rice genome. This is very interesting from an evolutionary standpoint. It is possible that MITEs were never derived from *OsHikkoshi* in rice, but were derived from the orthologous element in wheat. This allowed us to use the wheat *Hikkoshi* MITE to identify *OsHikkoshi* in the rice genome. The rice *Hikkoshi* MITEs may have originated from another related autonomous element, but such an element with complete features apparently no longer exists, at least in the Nipponbare genome. One possible explanation is that lesions in this autonomous element resulted in the loss of TIRs and subterminal regions, such that no traces of the rice *Hikkoshi*-like autonomous element were identified by the method employed here. A second possibility is that the parental autonomous element of the rice *Hikkoshi* MITEs has disappeared from the Nipponbare genome.

OsHikkoshi is also present in *indica* rice

When a BLAST search was conducted against whole genome shotgun (wgs) databases at NCBI, a sequence identical to *OsHikkoshi* was found in an *indica* rice genomic clone (AAAA02005857). However, this accession mapped to chromosome 2, while the accession containing *OsHikkoshi* in the *japonica* genome (AE017121) was located on chromosome 10. When we examined both loci reciprocally, we did not find *OsHikkoshi* or its excision footprints at the chromosome 10 locus of the *indica* genome, nor on the chromosome 2 locus of the *japonica* genome. This suggests that *OsHikkoshi* transposed into these loci after the evolutionary separation of *japonica* and *indica* rice.

Copy numbers of autonomous elements may influence the transposition of TEs. For example, copy numbers of *mPing* and *Pong* vary among cultivars of *japonica* and *indica* rice, and the increased copy number of *Pong* reflects the transposition activity of *mPing* (Jiang et al. 2003). It would be interesting to determine if the copy number of *OsHikkoshi* varies among rice cultivars, since this may influence the transposition activity of *Hikkoshi* family MITEs.

OsHikkoshi is a member of the *hAT* superfamily

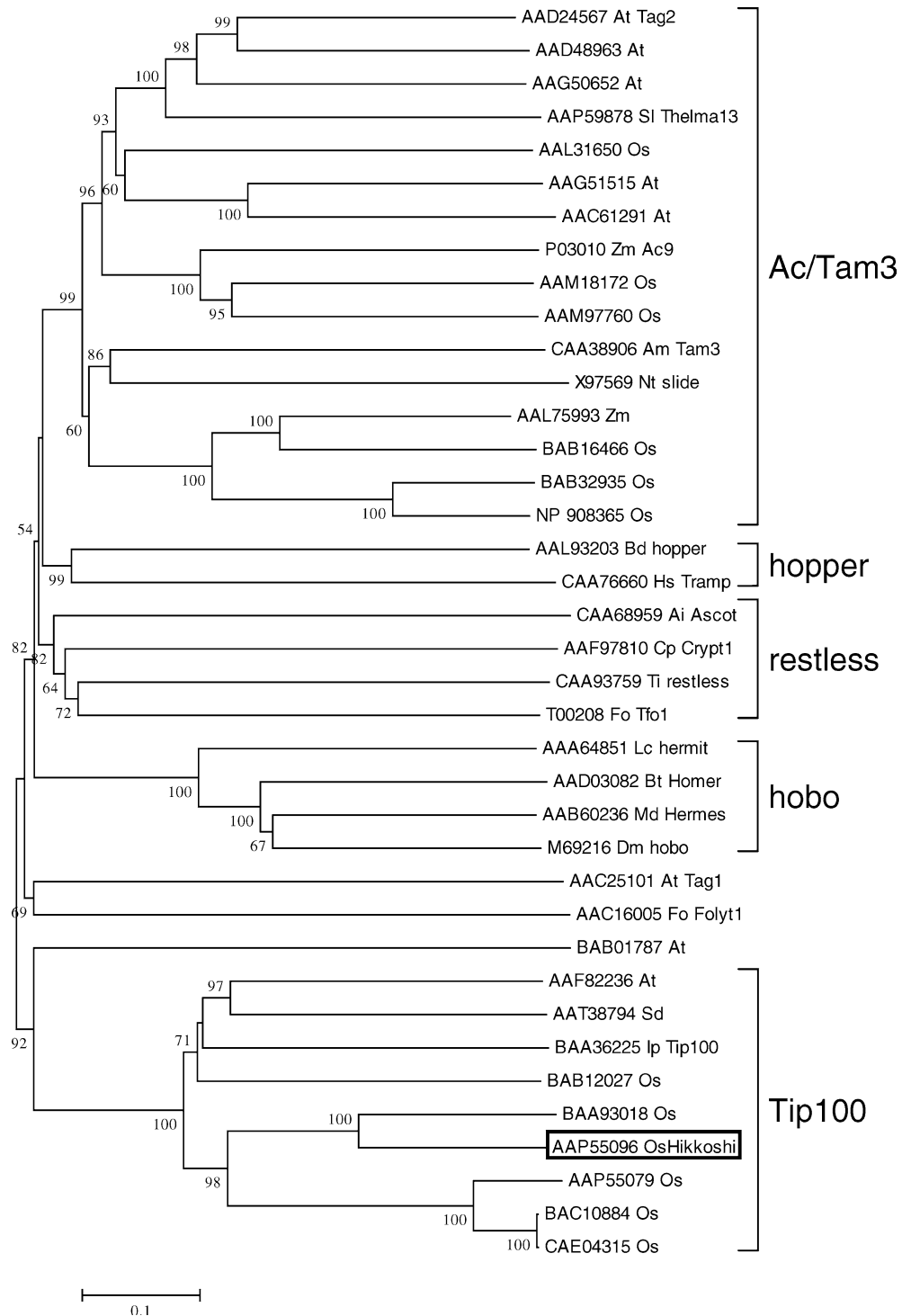
OsHikkoshi is 3424 bp in length and contains one ORF encoding 811 amino acids (Fig. 3b). The putative amino acid sequence includes two putative conserved domains, the ZnF_TTF and the hATC domains. ZnF_TTF is a zinc-finger domain found in TPases and transcription factors that is assumed to have DNA binding activity (Doerks et al. 2002). The hATC domain is a hAT dimerization domain (Essers et al. 2000) that is conserved among autonomous TEs of the *hAT* superfamily, such as *hobo*, *Ac*, *Tam3*, *Slide* and *Tag1* (Calvi et al. 1991; Hehl et al. 1991; Essers and Kunze 1995; Grappin et al. 1996). Like *hAT* superfamily MITEs, autonomous TEs of the *hAT* superfamily share similar TIRs flanked by 8-bp TSDs and, in addition, encode TPases with DNA binding domains and C-terminal hATC domains (Essers et al. 2000; Rubin et al. 2001). *OsHikkoshi* has all these

characteristics and clearly belongs to the *hAT* superfamily. Thus, both an autonomous TE and a number of MITEs belonging to the *hAT* superfamily have been found in the rice genome using this database search.

To further classify *OsHikkoshi*, a database search with the BLASTP program was conducted using the *OsHikkoshi* TPase as a query. Fourteen proteins, originating from *Arabidopsis thaliana*, *Solanum demissum*

and *Ipomoea purpurea* as well as rice (data not shown), showed high homology to the *OsHikkoshi* TPase (*E*-values less than e^{-150}). All except one of these proteins had ZnF_TTF and hATC domains. *Hikkoshi*-like TIRs and 8-bp TSDs flanked seven of the fourteen TPases (Table 2). One of these proteins was the TPase of *Tip100* (Fig. 3b), originally found in the *CHS-D* gene of *I. purpurea* (Habu et al. 1998). *Tip100* is categorized as a

Fig. 4 Phylogenetic tree of hAT TPases. Each sequence is identified by a GeneBank accession number, followed by two letters indicating the species of origin. At *A. thaliana*, Sl *S. latifolia*, Os *O. sativa*, Zm *Z. mays*, Am *A. majus*, Nt *N. tabacum*, Bd *B. dorsalis*, Hs *H. sapiens*, Ai *A. immersus*, Cp *C. parasitica*, Ti *T. inflatum*, Fo *F. oxysporum*, Lc *L. cuprina*, Bt *B. tryoni*, Md *M. domestica*, Dm *D. melanogaster*, Ip *I. purpurea*. Bootstrap values greater than 50% are shown



hAT superfamily TE, and is a member of a subgroup of similar TEs making up the *Tip100* family (Robertson 2002). TIR sequences of *OsHikkoshi* and *Tip100* are highly conserved; and *OsHikkoshi* appears to belong to the *Tip100* family or to a new family that is closely related to *Tip100*. The other six TPase encoding elements shown in Table 2 are also closely related to *Tip100* elements; however, none show as much similarity to the wheat *Hikkoshi* MITE as *OsHikkoshi* does. To examine the relationship of *OsHikkoshi* to other *hAT* TPases, a phylogenetic analysis was performed with 37 *hAT* TPases from plants, animals and fungi (Fig. 4). This analysis confirmed that the *OsHikkoshi* TPase belongs to the *Tip100* clade.

A TBLASTN search of a wheat EST database identified several clones (BQ904805, BJ216929, CA501395, BJ313779, CD925523, AL819648, BJ238316, BJ209493, CK163928, CA745598, AJ745420, CA664836, BQ166729, BQ235954, CD888663, BJ303488 and BQ802425) with high similarity (E -value $\leq e-05$) to the TPase of *OsHikkoshi* (data not shown). The sequence information from these clones should facilitate the isolation of autonomous *Hikkoshi* TEs from wheat. More than ten other MITEs belonging to several superfamilies have been identified in wheat (Bureau and Wessler 1994; Anderson et al. 2001, 2002; Hessler et al. 2002; Ramakrishna et al. 2002; Li et al. 2004). However, autonomous elements related to these MITEs have not been found. The comparative genomic approach employing computational analyses described here should also be useful in identifying autonomous TEs related to these wheat MITEs.

Acknowledgements We thank Dr. Patricia Vrinten for helpful discussions and useful editorial comments on the manuscript, Dr. Ken-ichi Tsutsumi and Hisakazu Hagami for assistance with Southern analysis, and Dr. Ryuji Ishikawa and Dr. Shigeru Saito for helpful information. This study was supported by a grant from the Japan Science and Technology Corporation.

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anderson OD, Hsia CC, Torres V (2001) The wheat γ -gliadin genes: characterization of ten new sequences and further understanding of γ -gliadin gene family structure. *Theor Appl Genet* 103:323–330
- Anderson OD, Larka L, Christoffers MJ, MaCue KF, Gustafson JP (2002) Comparison of orthologous and paralogous DNA flanking the wheat high molecular weight glutenin genes: sequence conservation and divergence, transposon distribution, and matrix-attachment regions. *Genome* 45:367–380
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. *Nucleic Acids Res* 32:D138–D141
- Besansky NJ, Mukabayire O, Bedell JA, Lusz H (1996) *Pegasus*, a small terminal inverted repeat transposable element found in the *white* gene of *Anopheles gambiae*. *Genetica* 98:119–129
- Bureau TE, Wessler SR (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4:1283–1294
- Bureau TE, Wessler SR (1994) *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6:907–916
- Bureau TE, Ronald PC, Wessler SR (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc Natl Acad Sci USA* 93:8524–8529
- Calvi BR, Hong TJ, Findley SD, Gelbart WM (1991) Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator and Tam3. *Cell* 66:465–471
- Doerks T, Copley RR, Schultz J, Ponting CP, Bork P (2002) Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* 12:47–56
- Essers L, Kunze R (1995) Transposable elements *Bg* (*Zea mays*) and *Tag1* (*Arabidopsis thaliana*) encode protein sequences with homology to *Ac*-like transposases. *Maize Genet Coop News Lett* 69:39–41
- Essers L, Adolphs RH, Kunze R (2000) A highly conserved domain of the maize *Activator* transposase is involved in dimerisation. *Plant Cell* 12: 211–223
- Feng Q, et al. (2002) Sequence and analysis of rice chromosome 4. *Nature* 420:316–320
- Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329–341
- Feschotte C, Swamy L, Wessler SR (2003) Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* miniature inverted repeat transposable elements (MITEs). *Genetics* 163:747–758
- Goff SA, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100
- Grappin P, Audeon C, Chupeau MC, Grandbastien MA (1996) Molecular and functional characterization of Slide, and *Ac*-like autonomous transposable element from tobacco. *Mol Genet* 252:386–389
- Habu Y, Hisatomi Y, Iida S (1998) Molecular characterization of the mutable *flaked* allele for flower variegation in the common morning glory. *Plant J* 16:371–376
- Hehl R, Nacken WK, Krause A, Saedler H, Sommer H (1991) Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveals homologies to the *Ac* element from maize. *Plant Mol Biol* 16:369–371
- Hessler TG, Thomson MJ, Benschler D, Nachit MM, Sorrells ME (2002) Association of a lipoxigenase locus, *Lpx-B1*, with variation in lipoxigenase activity in durum wheat seeds. *Crop Sci* 42:1695–1700
- Holyoake AJ, Kidwell MG (2003) *Vege* and *Mar*: two novel *hAT* MITE families from *Drosophila willistoni*. *Mol Biol Evol* 20:163–167
- Iwamoto M, Higo K (2003) *Tourist C* transposable elements are closely associated with genes expressed in flowers of rice (*Oryza sativa*). *Mol Genet Genomics* 268:771–778
- Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active transposon family in rice. *Nature* 421:163–167
- Kikuchi K, Terauchi K, Wada M, Hirano HY (2003) The plant MITE *mPing* is mobilized in anther culture. *Nature* 421:167–170
- Kong XY, Gu YQ, You FM, Dubcovsky J, Anderson OD (2004) Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. *Plant Mol Biol* 54:55–69
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Kunze R, Weil CF (2002) The *hAT* and CACTA superfamilies of plant transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington DC, pp 565–610

- Lepetit D, Pasquet S, Olive M, Thézé N, Thiébaud P (2000) Glider and Vision: two new families of miniature inverted-repeat transposable elements in *Xenopus laevis* genome. *Genetica* 108:163–169
- Li W, Wan Y, Liu Z, Liu K, Liu X, Li B, Li Z, Zhang X, Dong Y, Wang D (2004) Molecular characterization of HMW glutenin subunit allele *1Bx14*: further insights into the evolution of *Glu-B1-1* alleles in wheat and related species. *Theor Appl Genet* 109:1093–1104
- Morgan GT, Middleton KM (1990) Short interspersed repeats from *Xenopus* that contain multiple octamer motifs are related to known transposable elements. *Nucleic Acids Res* 18:5781–5786
- Nagano H, Kunii M, Azuma T, Kishima Y, Sano Y (2002) Characterization of the repetitive sequences in a 200-kb region around the rice *waxy* locus: diversity of transposable elements and presence of veiled repetitive sequences. *Genes Genet Syst* 77:69–79
- Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. *Nature* 421:170–172
- Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: analysis and visualization of genetic variation. *EMBNEW NEWS* 4:14
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85:2444–2448
- Ramakrishna W, Dubcovsky J, Park YJ, Busso C, Emberton J, SanMiguel P, Bennetzen JL (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. 162:1389–1400
- Robertson HM (2002) Evolution of DNA transposons in eukaryotes. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington DC, pp 1093–1110
- Rubin E, Lithwick G, Levy AA (2001) Structure and evolution of the hAT transposon superfamily. *Genetics* 158:949–957
- Saito M, Konda M, Vrinten P, Nakamura K, Nakamura T (2004) Molecular comparison of *waxy* null alleles in common wheat and identification of a unique null allele. *Theor Appl Genet* 108:1205–1211
- Sasaki T et al (2002) The genome sequence and structure of rice chromosome 1. *Nature* 420:312–316
- Schenke D, Sasabe M, Toyoda K, Inagaki Y, Shiraishi T, Ichinose Y (2003) Genomic structure of the *NtPDR1* gene, harboring the two miniature inverted-repeat transposable elements, *NtToy1* and *NtStowaway101*. *Genes Genet Syst* 78:409–418
- Smit AF, Riggs AD (1996) *Tiggers* and other DNA transposon fossils in the human genome. *Proc Natl Acad Sci USA* 93:1443–1448
- Surzycki SA, Belknap WR (1999) Characterization of repetitive DNA elements in *Arabidopsis*. *J Mol Evol* 48:684–691
- The Rice Chromosome 10 Sequencing Consortium (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300:1566–1569
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Tu Z (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* 98:1699–1704
- Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J* 26:307–316
- Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003a) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. *Plant Cell* 15:1186–1197
- Wicker T, Guyot R, Yahiaoui N, Keller B (2003b) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. *Plant Physiol* 132:52–63
- Yu J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). *Science* 296:79–92
- Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR (2001) *P* instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci USA* 98:12572–12577
- Zhang X, Jiang N, Feschotte C, Wessler SR (2004) *PIF*- and *Pong*-like transposable elements: distribution, evolution and relationship with *Tourist*-like miniature inverted-repeat transposable elements. *Genetics* 166:971–986