Mol Genet Genomics (2001) 266: 417–424
DOI 10.1007/s004380100530

ORIGINAL PAPER

G. Yang · J. Dong · M.B. Chandrasekharan · T.C. Hall

Kiddo, a new transposable element family closely associated with rice genes

Received: 27 February 2001 / Accepted: 26 April 2001 / Published online: 21 June 2001

Abstract The promoter region of the rice *ubiquitin2* (*rubq2*) gene was found to be polymorphic between *japonica* (T309) and *indica* (IR24) lines as the result of a 270-bp deletion in T309. A TTATA footprint in the T309 *rubq2* promoter suggested that an excision event had occurred, and inspection of the 270-bp region present in IR24 revealed that it had all the characteristics of a miniature inverted repeat transposable element (MITE). Database searches showed that this element is a member of a new MITE family, which we have named *Kiddo*. Thirty-five complete *Kiddo* sequences were identified in existing rice genomic sequence databases. They could be arranged into four groups, within-group sequence identity was over 90%, with 65–75% identity between groups. The high sequence similarity within a group indicates that some *Kiddo* members were recently mobile and may still be active. An additional 24 decayed *Kiddo* sequences were detected. Interestingly, ~80% of 18 *Kiddo* members from annotated accessions lie within 530 bp of a coding sequence. That ~40% of *Kiddo* members present in genic regions reside in introns suggests that *Kiddo* transposition entails the use of both DNA and RNA intermediates, and may provide some insight into the origins of individual groups. DNA blot analysis showed that *Kiddo* is a rice-specific element, although one sequence with limited (72%) similarity to *Kiddo* group A was detected as a wheat EST. *Kiddo* family members may represent new molecular and phylogenetic markers, as well as representing valuable materials for studying the molecular mechanisms of MITE transposition.

Communicated by G. P. Georgiev

G. Yang · J. Dong · M.B. Chandrasekharan · T.C. Hall (✉)
Institute of Developmental and Molecular Biology,
Texas A & M University, College Station,
TX 77843-3155, USA
E-mail: tim@idmb.tamu.edu
Tel.: +1-979-8457750
Fax: +1-979-8624098

Keywords Ubiquitin promoter · Intron · Transposition · Excision

Introduction

Following the discovery of transposable elements (TEs) in maize by Barbara McClintock (McClintock 1947), TEs have been found to be ubiquitous in biological organisms. DNA TEs are classified as transposons and elements that transpose through an RNA intermediate are classified as retrotransposons (Finnegan 1992). A typical transposon family has two types of members – the autonomous element which expresses a transposase, and non-autonomous elements which depend on that transposase for transposition. The DNA elements include terminal inverted repeats (TIRs) at both ends, while the internal sequences usually share substantial similarity among members of a given transposon family. Insertions result in target-site duplications (TSDs) in the genomic DNA immediately adjacent to the transposon. Excision of the element from genomic DNA brings the two TSDs together. A few nucleotides are sometimes removed at the insertion site by nucleases during the repair process (Bennetzen 2000). The excised transposon can then integrate into another genomic DNA location. The insertion event usually results in direct duplication of the target locus sequence.

In recent years, numerous small transposon-like elements have been found in many organisms. Collectively, these are called miniature inverted repeat transposable elements (MITEs) (Bureau et al. 1996). Each MITE family has a distinctive TIR and TSD, and a similar internal sequence; however, unlike *Ac*/*Ds* and *En*/*Spm*, the families do not appear to contain an autonomous element. Examples from plants include: *Tourist* (Bureau and Wessler 1992, 1994b), *Stowaway* (Bureau and Wessler 1994a), *Alien* (Pozueta-Romero et al. 1996), *Bigfoot* (Charrier et al. 1999), *Amy*/*LTP*, *p-Sine*, *Explorer*, *Gaijin*, *Castaway*, *Ditto*, *Wanderer* (Bureau et al. 1996), *Emigrant* (Casacuberta et al. 1998),

Krispie, *Snap*, *Crackle*, *Pop*, *Snabo-1*, *Snabo-2*, *Snabo-4*, *Truncator* (Mao et al. 2000), *Hbr*, *mPIF*, *Olo* (Zhang et al. 2000). It is likely that additional MITE families remain to be discovered. Here we report a new MITE family, which we have named *Kiddo*.

MITEs have the characteristics of both DNA and RNA elements (Wessler et al. 1995). They contain typical DNA element structures, such as TIRs and TSDs. However, their very high copy number and lack of evidence for excision is more characteristic of retrotransposons. Classification of MITEs as DNA elements depends on the existence of DNA intermediates during transposition or evidence for their excision from genomic DNA. Unfortunately, no direct evidence for transposition has been reported for MITEs and the only excision footprint observed thus far is in *Hbr* (Zhang et al. 2000). Here we report evidence for the excision of a *Kiddo* element from rice genomic DNA.

Some MITE families in grasses, for example, *Tourist* and *Hbr*, were originally found in maize. MITE family members in other grasses such as rice, wheat and sorghum have been found either by using the maize MITE sequences as primers to probe their genomic DNA, or by the use of maize MITE sequences as queries for database searches (Bureau et al. 1996; Oosumi et al. 1995). Many MITE families originally found in maize also contain members from other grasses (Bureau and Wessler 1994b; Mao et al. 2000; Zhang et al. 2000). However, this approach will not reveal MITE families, such as *Kiddo*, that exist only in rice, and, for rice-specific MITEs, the study of individual gene polymorphisms remains an important way to discover new families (Bureau et al. 1996).

MITEs are useful in systematics and are potentially useful as molecular markers because several are preferentially associated with genes (Bureau and Wessler 1994a, 1994b; Zhang et al. 2000). For example, *Hbr* has been successfully used as a molecular marker in maize because of its association with genic regions and the sequence homogeneity that exists within this family (Casa et al. 2000). The *Kiddo* family, which exists in a high copy number, also appears to have value as a new molecular marker since each subgroup shows over 90% sequence similarity and ~80% of the 18 family members identified thus far from annotated sequences lie less than 530 bp from cDNA sequences or from introns.

Materials and methods

PCR amplification of the *rubq2* promoter

Genomic DNA of the rice lines T309 and IR24 (kindly supplied by Anna M. McClung, Texas Agricultural Experimental Station, Beaumont, Tex.) was extracted using a hexadecyltrimethyl ammonium bromide method (Taylor and Powell 1982). DNA was digested with *Hin*dIII to facilitate PCR amplification of the desired product, as there is no *Hin*dIII site in the target region. PCRs were conducted at various temperatures between 45° and 65°C with cloned *Pfu* (Stratagene) or *Taq* DNA polymerase (Promega, Madison, Wis.). The primers used were: 5′-AAGCTTACGGAAGGAAAACAAATTCGG-3′ and 5′-TCTAGATGCGAGG A GAGGAGATGAG-3′.

Cloning and sequencing of the PCR products

PCR products were cloned into pPCR-ScriptJ™ Amp SK (+) and transformed into Epicurian Coli® XL10-Gold® ultracompetent cells (Stratagene). White colonies were selected on X-Gal LB plates with IPTG. Positive transformants were identified by agarose gel electrophoresis after digestion of isolated plasmids with *Sac*II. Automated sequencing of the cloned PCR fragments was done by the Texas A & M Gene Technologies Laboratory, using T3 and T7 promoter-specific primers.

DNA blot analysis

Aliquots (2 μg) of genomic DNA from *Arabidopsis*, maize, tobacco, rice T309, rice IR24, wheat and *Camptotheca acuminata* were digested with *Hin*dIII overnight at 37°C. After electrophoretic fractionation of DNA on 1% agarose gel for 12–15 hours at a constant 23 V, the DNA was transferred to Hybond™-N+ nylon membrane (Amersham, Piscataway, N.J.). Genomic DNA blot analysis was performed as described by Buchholz et al. (1998). [^{32}P]dCTP-labeled probes were made using a DECAprime™ II DNA labeling kit (Ambion, Austin, Tex.). Membranes were washed with 2×SSC [1×SSC = 0.15 M sodium chloride, 0.015 M sodium citrate (pH 7)]/0.1% SDS at 65°C for 1 h (low stringency), or with 0.3×SSC/0.1% SDS at 65°C for 1 h (moderate stringency).

Data mining and alignment

Initially, a BLASTN search was undertaken against GenBank, EMBL, DDBJ, PDB, EST, STS, GSS, and HTGS databases (current on December 23, 2000; National Center for Biotechnology Information, Bethesda, Md.), using the 270-bp deletion sequence in the IR24 *rubq2* promoter region as a query. The retrieved sequences were then used as queries to carry out BLASTN and TBLASTX searches in the databases. Protein databases were used to retrieve any possible translation product and BLASTX searches were used to study the putative coding capacity of the retrieved elements. All retrieved sequences with expectation values lower than 1×10^{-3} and with a lengths exceeding 150 bp were studied further. After elimination of duplicate sequences, complete sequences were aligned using Vector NTI suite 6.0 AlignX (InforMax, Bethesda, Md.) with a gap penalty of 5 and 15 and gap-extension penalties of 1 and 6.66 for pairwise alignment and multiple alignment, respectively. Alignments were visualized using BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). A phylogenetic tree was produced using Vector NTI 6.0 AlignX. Folding of DNA fragments was carried out using M-fold (http://bioinfo.math.rpi.edu/~mfold/dna/form1.cgi).

Results

Discovery of *Kiddo*

Genomic DNA from *O. sativa* (L.) cv. Taipei T309 was digested with *Hin*dIII, and used as a template for PCR amplification of the rice *ubiquitin2* (*rubq2*) promoter region. A single discrete product was obtained, but was smaller (684 bp) than expected (954 bp) from the database sequence (GenBank Accession No. AF184280). The amplified fragment was cloned into PCRscript-Amp SK (+) and sequencing confirmed that, except for a 270-bp deletion, the sequence was identical to the published

sequence of *rubq2* (Wang et al. 2000). The cloned fragment was confirmed to be derived from the *rubq2* gene by genomic DNA blot analysis (data not shown).

Since the original *rubq2* sequence in GenBank was obtained from a BAC clone derived from the rice line IR24, we obtained that line and repeated the PCR amplification using the same primers as used for T309. Cloning and sequencing of the amplified product confirmed it to be identical to the expected 954 bp of *rubq2*. New genomic DNA extracts were made and independent replicate experiments confirmed that PCR products of dissimilar sizes were obtained from T309 and IR24 using *Taq* DNA polymerase (Fig. 1A), indicating an insertion of 270 bp in *rubq2* from IR24 (or a deletion in T309). Interestingly, *Pfu* DNA polymerase failed to amplify the *rubq2* fragment from IR24 under the conditions used (Fig. 1B), suggesting that the insertion might contain a secondary structure inimical to the polymerase. The location of the insertion in the *rubq2* promoter of IR24 is shown in Fig. 1C.

Inspection of the sequence of the 270-bp region present in *rubq2* from IR24 revealed that its ends constituted 14-bp terminal inverted repeats. To examine the possibility that the 270-bp insertion represented a TE, it was used as a query sequence to search the database (see Materials and methods). Twelve rice sequences with high (over 90%) similarity were retrieved that all had MITE characteristics – small size, TIR and TSD (Fig. 2A). All were 269 bp in length and, since their internal sequences had no detectable similarity to those of known TEs, we classified them as a new MITE family named *Kiddo*. This set of sequences (*Kiddo* group A) has

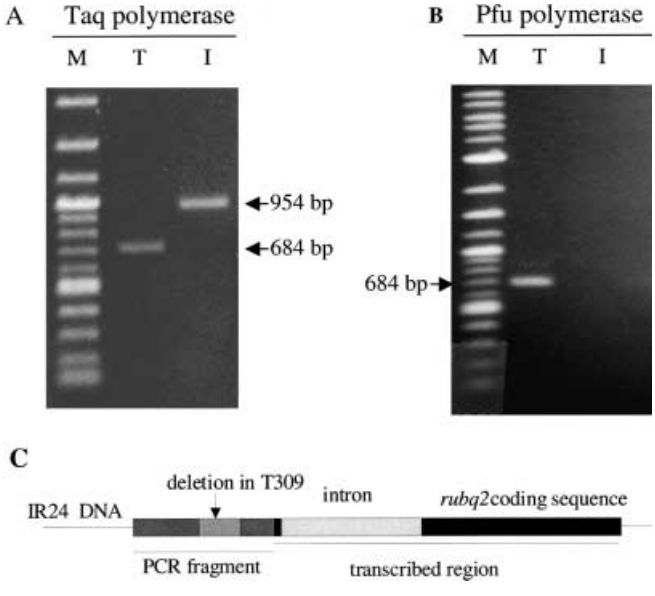

Fig. 1A–C PCR amplification of the *rubq2* promoter. **A** Using *Taq* DNA polymerase (**A**) and *Hin*dIII-digested genomic DNA, a 684-bp fragment was PCR-amplified from T309 (T) and a 954 bp fragment was amplified from IR24 (I). **B** PCR using *Pfu* DNA polymerase amplifies only the 684-bp fragment from T309 DNA. **C** Diagram of IR24 *rubq2* showing the location of the PCR-amplified region containing the *Kiddo* element

the consensus TIR GGGGCTGTTTGGTT, with two mismatches at the 4th and 5th nucleotides. No polyA/polyT elements were found in the subterminal regions. The TSDs consisted of 3 nt. The first was typically T, the second was usually T or A, and the third was preferentially T or A. Some 10 bp of A/T-rich sequence flanked the TSDs. Four copies of the motif TTGCCA and one plant G-box factor binding site (TGACACGTGGG; Tfsitescan program expectation value 1.28e-03) were typically present within the internal sequence. An additional TE-like sequence flanked by TTTTTTGA and TCAAATTT was found nested inside the group A sequences (Fig. 2A). As shown in Fig. 2B for *Kiddo* in IR24 *rubq2*, all of the sequences in this group have the potential to form hairpin structures. This could result in the formation of cruciform structures in rice genomic DNA; since such structures are known to be capable of affecting transcriptional regulation (Wadkins 2000), the presence of *Kiddo* MITEs may affect the expression of adjacent genes, such as the IR24 *rubq2* gene.

Excision of *Kiddo* from the T309 *rubq2* promoter region

T309 belongs to the *japonica* subspecies of rice, while IR24 is a member of the *indica* subspecies (Jiang et al. 1995). These subspecies are thought to have started to diverge some 2–3 million years ago (Huke and Huke 1990). The question arises whether the *Kiddo* element inserted into IR24 after divergence of the subspecies or was present before divergence and left the locus in T309 due to transposition following divergence. In the case of a de novo insertion, one can postulate that the target sequence was TTA↓GA prior to insertion of a TE and TTA*e*TTAGA after insertion (where *e* represents the TE insert) (Fig. 3). In the case of a TE excision from this insertion, one would expect TTA↓TTAGA after precise excision or TTA↓TAGA if a single nucleotide (T) was lost at the right-border TSD (Saedler and Nevers 1985). Inspection of the rice lines shows that the IR24 sequence TA*e*TTAGA is in agreement with retention of a TE insertion and that the T309 sequence TTA↓TAGA is that expected after loss of the TE; i.e. the TE (*Kiddo*) was present prior to divergence. An alternative possibility is that the target sequence was TTA↓TAGA and that the T309 sequence reflects lack of any TE insertion. However, it is then difficult to see how insertion of a TE into this sequence could occur (as evidently did occur in IR24) since the expected sequence would be TTA*e*TTATAGA, which is not observed.

Four groups of *Kiddo*-related elements can be found in rice

After several rounds of database searching, 36 *Kiddo*-related sequences were retrieved that included both TIRs. These were sorted (see Materials and methods) into four

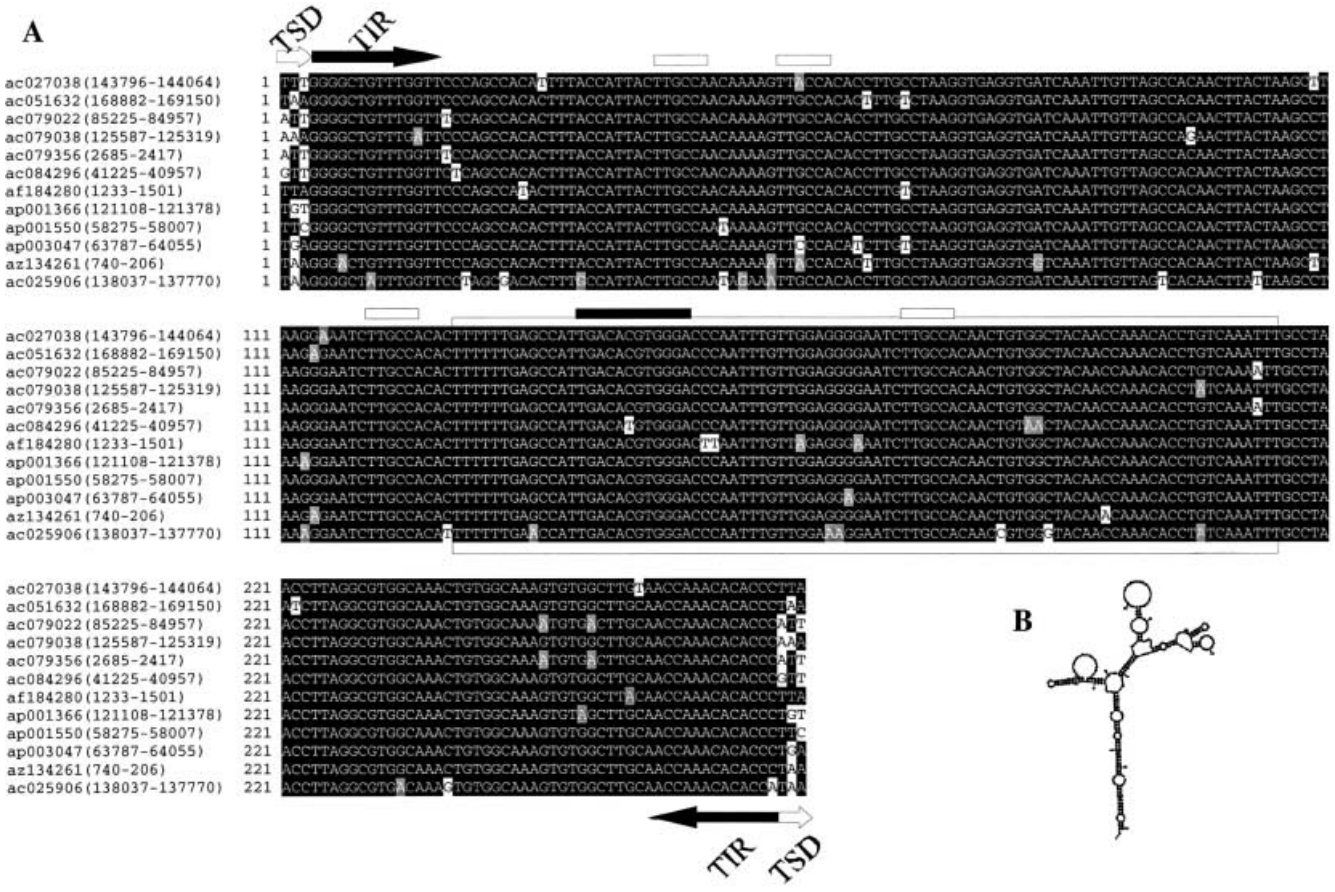

Fig. 2A Sequence alignment of twelve genomic members of *Kiddo* group A. The sequences were aligned using Vector NTI 6.0 AlignX and visualized with BOXSHADE. The *filled arrows* refer to the TIRs and *empty arrows* denote TSD sequences. The *filled bar* indicates a G box factor binding site. The *open bars* denote TTGCCA repeats. The *boxed region* represents sequences flanked by TTTTTTGA and TCAAATTT. **B** Folding of the 270-bp insertion fragment from IR24 *rubq2* promoter as predicted by M-fold (see Materials and methods). $\Delta G = -37.6$ kcal/mol at 37°C

groups (Fig. 4). They all had MITE characteristics, with internal sequences that are flanked by AT-rich microregions (data not shown). However, individual groups had slightly different consensus TIRs and TSDs. The derived phylogenetic tree (Fig. 4) indicates that the *Kiddo* family diverged from a common ancestor. Presumably, during evolution, some sequence changes inactivated the ability to transpose while others had little or a positive effects on transposition that permitted their propagation throughout rice genomes. Groups A, B, C and D comprise multiple copies of highly similar sequences, and are thus likely to be still active. Within each of the four groups, sequences are more than 90% identical. Similarity between groups ranges from 65 to 75%.

Kiddo is a rice-specific element

In addition to the 35 complete *Kiddo* sequences, 24 incomplete sequences were retrieved, with 75–93% similarity to queries (see Materials and methods). That 58 out of 59 retrieved sequences came from rice suggested that the *Kiddo* MITE family was essentially restricted to the rice genome. One sequence from wheat (*Kiddo-ta1*) had an overall similarity (~72%) to group A *Kiddo* elements, with ~56% similarity within a region flanked by TTTTTTGA and TCAAATTT sequences and ~83% outside this region. That *Kiddo* is essentially confined to the rice genome was supported by the results of genomic DNA blot analysis. When genomic DNAs from *Arabidopsis*, tobacco, maize, wheat, a tree (*C. acuminata*), rice IR24 and rice T309 were probed with the *Kiddo-os11* PCR product, only rice DNAs hybridized (Fig. 5). This was true at both low and moderate stringency wash conditions.

The question arises as to why the *Kiddo* element has left the *rubq2* promoter in T309 but not IR24. It has been proposed that certain genomic contexts provide more favorable environments than others for transposition (Kalendar et al. 2000). Comparison of the

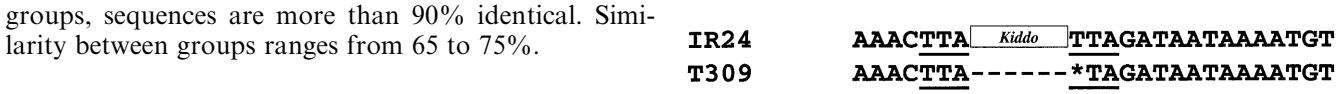

Fig. 3 *Kiddo* was lost from the T309 *rubq2* promoter. The *Kiddo* element in the *rubq2* promoter in IR24 is flanked by TSDs (*underlined*). The element is absent from this locus in T309 plants (*dashed line*) and a T residue has been lost (*asterisk*) from the right-border TSD

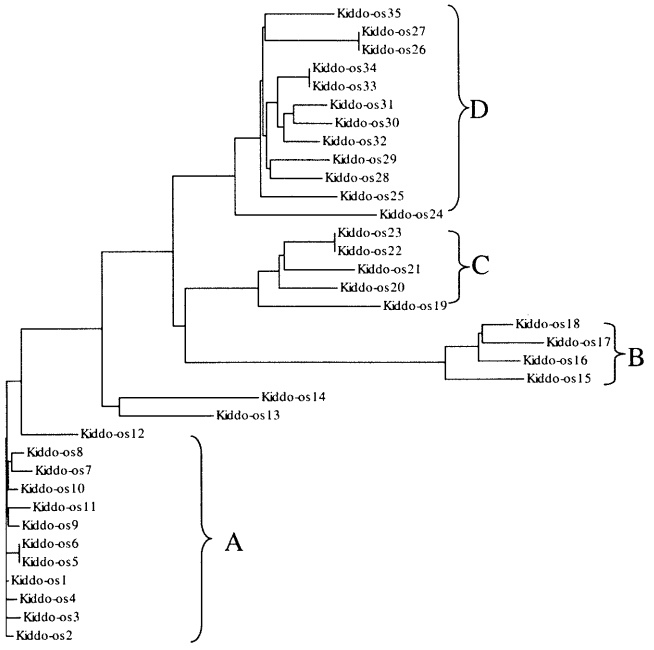

Fig. 4 Phylogenetic tree of *Kiddo* sequences. The tree was generated using Vector NTI AlignX. Numbering for family members corresponds to that in Table 1. Groups A–D comprise elements that are over 90% similar in sequence

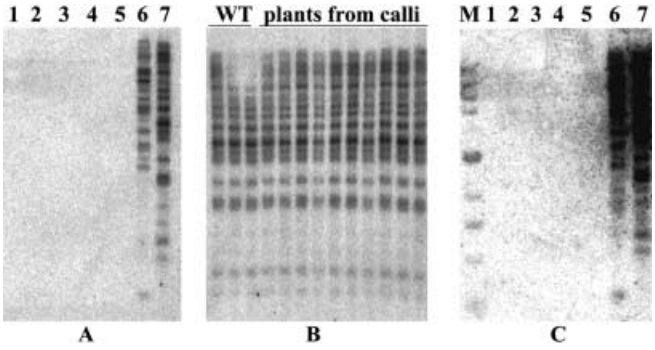

Fig. 5A–C *Kiddo* is prevalent in the rice genome. **A** Various genomic DNAs were probed with the *Kiddo-os11* sequence. **B** Genomic DNAs from three T309 plants (WT) and ten T309 plants regenerated from independent calli were probed with the *Kiddo-os11* sequence. **C** Genomic DNAs isolated from various plant species were probed with a combination of *Kiddo* group A, B, C and D sequences. Lanes: 1, *Camptotheca acuminata*; 2, tobacco; 3, *Arabidopsis*; 4, wheat; 5, maize; 6, rice IR24; 7, rice T309

hybridization profiles for T309 and IR24 probed with *Kiddo-os11* shows much greater complexity for T309 (Fig. 5A), possibly suggesting a more favorable climate for transposition exists in this line. However, no transposition was detected when 10 plants were regenerated from independent T309 calli (Fig. 5B), suggesting that tissue culture does not stimulate migration of *Kiddo*. A mixture of group A, B, C and D fragments was also used to probe the various genomic DNAs. That only rice genomic DNAs showed hybridization further supports the contention that *Kiddo* is a rice-specific MITE (Fig. 5C). The intense banding profile seen with these combined probes reveals that the *Kiddo* family exists in high copy numbers. The total copy number of *Kiddo* in the rice genome is estimated to be about 3000, based on our identification of 58 sequences from current databases that represent only ~2% of the rice genome.

Kiddo elements are closely associated with genes

Table 1 lists 39 members of the rice *Kiddo* family and one putative member from wheat. An additional 19 candidate members with incomplete TIRs were found in non-annotated sequences (data not shown). Of the 59 *Kiddo*-like sequences identified from public databases, 18 were in annotated sequences. The other 41 sequences were not annotated, and it is possible that some or all of these are present in genic regions. The distribution of the *Kiddo* elements in genomic DNA was examined with reference to coding sequences (CDS). *Kiddo-os37* and *Kiddo-os25* were located 4168 bp and 2789 bp, respectively, from a CDS. *Kiddo-os32* and *Kiddo-os38* were at about 1 kb from a CDS. Six *Kiddo* elements resided within 530 bp (5′ or 3′) of a CDS and seven were in intron regions. The related sequence from wheat was an EST sequence. The fact that 14 of the 18 (~80%) sequences were within 530 bp of CDSs (Table 1), or in introns, underscores the close association of *Kiddo* elements with genes. This property suggests that *Kiddo* family members are suitable new molecular markers for genic regions in rice.

Discussion

Origin and amplification of *Kiddo*

MITEs occasionally share similar TIRs with autonomous DNA transposons and thus are thought to have derived from DNA transposons (Morgan 1995; Unsal and Morgan 1995; Yeadon and Catcheside 1995; Feschotte and Mouches 2000). However, MITEs usually occur in very high copy numbers (Morgan 1995; Unsal and Morgan 1995; Wessler et al. 1995; Zhang et al. 2000), which can not be achieved by classical DNA transposition. An alternative origin of MITEs has been suggested that involves aberrant DNA replication events when DNA polymerases encounter palindromic sequences as templates. In this model (Izsvak et al. 1999), the 3′ region of a nascent DNA strand may fold back, allowing DNA synthesis to reinitiate using the nascent DNA strand as template. A stem loop byproduct (the *Angel* MITE) may result from this aberrant replication. After its excision from genomic DNA, the stem loop can then integrate into other genomic DNA locations with the help of recombinases, providing new sites for amplification of the MITE.

The distribution of *Kiddo* members with respect to genic regions may provide insight into the amplification of the family. Of 18 *Kiddo* members in annotated

Table 1 Genomic distribution of *Kiddo*

Kiddo element[a]	Accession No.[b]	Genic insertion[c]	Position
os1	ap001550 (58275–58007)	Yes	523 bp from CDS 3'
os2	ap001366 (121108–121378)	Yes	Intron
os3[d]	ac084296 (41225–40957)	?	?
os4[d]	ac079038 (125587–125319)	?	?
os5[d]	ac079356 (2685–2417)	?	?
os6[d]	ac079022 (85225–84957)	?	?
os7	az134261 (740–206)	?	?
os8H	ac027038 (143796–144064)	?	?
os9	ap003047 (63787–64055)	?	?
os10[d]	ac051632 (168882–169150)	?	?
os11	af184280 (1233–1501)	Yes	210 bp from TATA 5′
os12[d]	ac025906 (138037–137770)	?	?
os13	aj245900 (62716–62965)	Yes	Intron
os14	aq157268 (196–389)	?	?
os15[d]	ac083943 (13936–13662)	?	?
os16	ap003048 (121712–121988)	?	?
os17	aq157046 (280–4)	?	?
os18	ap002743 (32729–33008)	Yes	379 bp from CDS 5′
os19[d]	ac084282 (118802–119079)	?	?
os20	aq795953 (109–386)	?	?
os21	aq689856 (45–319)	?	?
os22[d]	ac027037 (118552–118828)	?	?
os23[d]	ac018929 (9993–10269)	?	?
os24	ap002816 (24995–24724)	Yes	Intron
os25	ac082644 (108393–108633)	No	2789 bp from CDS 3′
os26	ap002864 (2288–2559)	?	?
os27	ab023482 (148318–148589)	Yes	232 bp from CDS 3′
os28	ap000836 (148097–147854)	Yes	Intron
os29	al442114 (44482–44203)	Yes	Intron
os30	aq273730 (524–254)	?	?
os31[d]	ac079852 (205709–205988)	?	?
os32	ac007858 (28501–28777)	Yes	1023 bp from CDS 5′
os33[d]	ac079029 (125597–125876)	?	?
os34	ac051634 (134585–134306)	Yes	126 bp from CDS 5′
os35	ab026295 (26954–27202)	Yes	367 bp from CDS 5′
ta1	be517419 (325–49)	Yes	cDNA
os36[e]	af128457 (66418–66286)	Yes	Intron
os37[e]	ac069145 (33249–32753)	No	4168 bp from CDS 3′
os38[e]	ap002522 (113573–113331)	Yes	1203 bp from CDS 3′
os39[e]	af119222 (73595–73723)	Yes	Intron

[a]os, *Oryza sativa*; ta, *Triticum aestivum*
[b]Numbers in *parentheses* indicate the position of the element in the cited accession
[c]Based on distance from a TATA box or CDS; elements less than 2 kb from a CDS are considered genic; os11 is from IR24 *rubq2*. ?, unknown (genes have not yet been identified for these annotated sequences)
[d]HTGS sequences, positions in the accessions are subject to change
[e]Incomplete elements found in annotated sequences

sequences, seven (40%) are in introns, six (30%) are less than 530 bp from a CDS and four (20%) lie between 1 and 5 kb from a CDS. Although this distribution shows clustering around genic regions, it is interesting that no MITEs are in coding sequences. This organization may indicate that the origin of the groups within the *Kiddo* family resides within introns. Such an arrangement could yield a very large number of *Kiddo* copies as a result of transcription and subsequent excision by splicing. Reverse-transcription of a proportion of the excised *Kiddo* RNA elements into DNA might confer transposition capability, permitting integration into nearby genomic DNA locations. Transposition into introns of other genes could lead to further propagation of *Kiddo*. This scenario implies the possibility that RNA intermediates might be involved in transposition of *Kiddo*, in addition to, or instead of, DNA intermediates.

Excision and transposition of *Kiddo-os11*

Since the widespread distribution of MITEs implies an ability to transpose, it is curious that movement has not been observed. If autonomous elements capable of supporting MITE transposition exist, it is likely that they are present on elements separate from the MITEs, since MITEs (including *Kiddo*) do not have extensive ORFs (Bureau and Wessler 1992). It is conceivable that a transposase is provided in trans from other genes resident in the organism.

The excision of *Kiddo* from the rice T309 *rubq2* promoter appears to be a net excision, with modest damage at the first nucleotide (T) of the right side TSD. Degradation of the overhanging 5′-ends of the TSD has been observed for other DNA transposons (Saedler and Nevers 1985). Previous evidence of excision footprints

for *Hbr* (Zhang et al. 2000), together with the present evidence for *Kiddo* (Fig. 3), suggests that MITEs can excise from genomic DNA. This supports the concept that MITEs are similar to DNA elements (Bureau and Wessler 1994b). However, caution is advisable in accepting footprints as evidence for MITE transposition, as retrotransposons can also excise from genomic DNA at a very low frequency (Bennetzen 2000). Investigation of the ability of MITEs to transpose has been hindered by the fact that families with low sequence similarity may contain decayed MITE members that have lost their mobility. In contrast, highly homogeneous MITE families such as *Hbr* and *Kiddo* may well still be functional and may, therefore, be useful candidates for solving the mystery of MITE transposition.

Significance of the TIR sequence

TIRs are very important for transposons and are thought to serve as transposase binding sites (Becker and Kunze 1997), perhaps accounting for their high conservation in a given transposon family. The same TIR can be found in various MITE families from different organisms. The TIR sequence GGGGNT GTTTGGTT, which is present in *Tourist-D* and *Hbr* (Bureau and Wessler 1994b; Zhang et al. 2000), is also found in *Kiddo* (Fig. 2). There are two explanations for the presence of this TIR in rice, maize and wheat. One is that MITEs containing this TIR are derived from a common ancestor, and that the internal sequences have undergone massive mutation so that they no longer show recognizable similarity. This explanation assumes that the internal sequences are not functionally necessary. An alternative explanation is that sequences bearing this TIR are especially susceptible to aberrant replication events (Izsvak et al. 1999), leading to the de novo creation of MITE families bearing the same TIR, but with no detectable similarity in their internal sequences. In both models, the TIR is considered to be very important for MITE propagation. Studies on protein factor binding interactions are needed to reveal the role of this TIR in the transposition of *Tourist-D, Hbr* and *Kiddo*.

Involvement in the regulation of gene expression

The close association of MITEs with plant genes may indicate their involvement in the evolution of these genes (Wessler et al. 1995; Bureau et al. 1996). The close proximity of *Kiddo* members to CDSs suggests that the insertion of these elements could probably modify transcriptional, splicing or translational regulation of the genes. The IR24 *rubq2* promoter that contains the *Kiddo* insertion has been shown to drive high levels of reporter gene expression in transient assays (Wang et al. 2000), and it is possible that the G-box present in *Kiddo-os11* augments transcriptional activity. Therefore, it will be informative to compare the activity of *rubq2* promoters from IR24 and T309 in stably transformed rice plants to determine if the presence of the *Kiddo* insertion increases or decreases promoter strength.

Acknowledgements We thank the Texas Advanced Technology Program (Grant 010366-0092-1999) for financial support, and Yiming Jiang, Magda Cervera, Sophie Fernandez and Tao Wang for helpful discussions and providing materials.

References

Becker HA, Kunze R (1997) Maize *Activator* transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. Mol Gen Genet 254:219–230

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251–269

Buchholz WG, Connell JP, Kumpatla SP, Hall TC (1998) Molecular analysis of transgenic rice. Methods Mol Biol 81:397–415

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 (1994a) *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, Wessler SR (1994b) Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. Proc Natl Acad Sci USA 91:1411–1415

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

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) The MITE family *heartbreaker* (*Hbr*): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Casacuberta E, Casacuberta JM, Puigdomenech P, Monfort A (1998) Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. Plant J 16: 79–85

Charrier B, Foucher F, Kondorosi E, d'Aubenton-Carafa Y, Thermes C, Kondorosi A, Ratet P (1999) *Bigfoot*. a new family of MITE elements characterized from the *Medicago* genus. Plant J 18:431–441

Feschotte C, Mouches C (2000) Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. Mol Biol Evol 17:730–737

Finnegan DJ (1992) Transposable elements. Curr Opin Genet Dev 2:861–867

Huke RE, Huke EH (eds) (1990) Rice: then and now. International Rice Research Institute, Manila, Philippines

Izsvak Z, Ivics Z, Shimoda N, Mohn D, Okamoto H, Hackett PB (1999) Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J Mol Evol 48:13–21

Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci USA 92:4487–4491

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607

Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA,

Wing RA (2000) Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res 10:982–990

McClintock B (1947) Cytogenetic studies of *maize* and *Neurospora*. Carnegie Inst Wash Yearbook 46:146–152

Morgan GT (1995) Identification in the human genome of mobile elements spread by DNA- mediated transposition. J Mol Biol 254:1–5

Oosumi T, Garlick B, Belknap WR (1995) Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. Proc Natl Acad Sci USA 92:8886–8890

Pozueta-Romero J, Houlne G, Schantz R (1996) Nonautonomous inverted repeat *Alien* transposable elements are associated with genes of both monocotyledonous and dicotyledonous plants. Gene 171:147–153

Saedler H, Nevers P (1985) Transposition in plants: a molecular model. EMBO J 4:585–590

Taylor B, Powell A (1982) Isolation of plant DNA and RNA. Focus 4:4–6

Unsal K, Morgan GT (1995) A novel group of families of short interspersed repetitive elements (*SINE*s) in *Xenopus*: evidence of a specific target site for DNA-mediated transposition of inverted-repeat *SINE*s. J Mol Biol 248:812–823

Wadkins RM (2000) Targeting DNA secondary structures. Curr Med Chem 7:1–15

Wang J, Jiang J, Oard JH (2000) Structure, exprssion and promoter activity of two polyubiquitin genes from rice (*Oryza sativa* L.). Plant Sci 156:201–211

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Yeadon PJ, Catcheside DE (1995) *Guest*: a 98 bp inverted repeat transposable element in *Neurospora crassa*. Mol Gen Genet 247:105–109

Zhang Q, Arbuckle J, Wessler SR (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. Proc Natl Acad Sci USA 97:1160–1165