Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 8524–8529, August 1996
Genetics

A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes

(transposons/SINEs/LINEs/MITEs/repetitive DNA)

THOMAS E. BUREAU*†, PAM C. RONALD‡, AND SUSAN R. WESSLER*

*Department of Genetics, University of Georgia, Athens, GA 30602; and ‡Department of Plant Pathology, University of California, Davis, CA 95616

Communicated by Steven D. Tanksley, Cornell University, Ithaca, NY, March 1, 1996 (received for review December 1, 1995)

ABSTRACT **Several recent reports indicate that mobile elements are frequently found in and flanking many wild-type plant genes. To determine the extent of this association, we performed computer-based systematic searches to identify mobile elements in the genes of two "model" plants, *Oryza sativa* (domesticated rice) and *Arabidopsis thaliana*. Whereas 32 common sequences belonging to nine putative mobile element families were found in the noncoding regions of rice genes, none were found in Arabidopsis genes. Five of the nine families (*Gaijin*, *Castaway*, *Ditto*, *Wanderer*, and *Explorer*) are first described in this report, while the other four were described previously (*Tourist*, *Stowaway*, p-SINE1, and Amy/LTP). Sequence similarity, structural similarity, and documentation of past mobility strongly suggests that many of the rice common sequences are bona fide mobile elements. Members of four of the new rice mobile element families are similar in some respects to members of the previously identified inverted-repeat element families, *Tourist* and *Stowaway*. Together these elements are the most prevalent type of transposons found in the rice genes surveyed and form a unique collection of inverted-repeat transposons we refer to as miniature inverted-repeat transposable elements or MITEs. The sequence and structure of MITEs are clearly distinct from short or long interspersed nuclear elements (SINEs or LINEs), the most common transposable elements associated with mammalian nuclear genes. Mobile elements, therefore, are associated with both animal and plant genes, but the identity of these elements is strikingly different.**

Transposable elements are an integral component of most, if not all, genomes. In fact, the majority of interspersed repetitive DNA may be composed of transposons (1). Transposable elements are not merely passive components of genomes as they may play an important role in the evolution of developmental processes (2). The plethora of mutant phenotypes generated by transposon insertions into or near nuclear genes powerfully illustrates this potential.

The actual role of mobile elements in normal or wild-type gene evolution, however, is much less clear. Ancient and recent insertions of transposons are evident from the sometimes large number of elements in and flanking mammalian genes (3, 4). A few of these elements have been documented as providing cis-factors influencing the expression of nearby genes (5). Short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs) are the predominant mobile elements in association with mammalian genes and belong to the general class of transposons that move via an RNA intermediate (3, 4). These elements have distinctive structural characteristics and thus are easily distinguished from members of the other general class of transposons that move via a DNA intermediate (6).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Recently several studies have shown that wild-type plant genes also harbor mobile elements (7–13). Retrotransposons of the *copia* class are ubiquitous in plant genomes and have been found flanking wild-type plant genes (11, 14–17). SINEs and LINEs are also present in plant genomes but are less frequently observed near plant gene sequences (8, 13, 18, 19; S. White, personal communication). In contrast, the unusual inverted-repeat elements, *Tourist* and *Stowaway*, have been identified in association with over a hundred plant gene sequences (7, 9, 10). In this study, a comprehensive survey of the wild-type gene sequences from two plants, *Oryza sativa* (domesticated rice) and *Arabidopsis thaliana*, was conducted to ascertain the type and frequency of mobile elements associated with their genes.

MATERIALS AND METHODS

Computer Analysis. Rice and Arabidopsis gene sequences were compiled from the GenBank (version 78.0) and EMBL (version 35.0) databases using the program STRINGS as part of the University of Wisconsin Genetics Computer Group (UWGCG) (20) program suite (version 7.0) accessed through the Biological Sciences Computer Resource (University of Georgia, Athens). Rice, Oryza, sativa, Arabidopsis, and thaliana were used as keywords. Expressed sequence tags, sequenced tagged sites, cDNA sequences, and cytoplasmic sequences were not used as queries in this study. After duplicate sequences were eliminated, a list of 105 rice and 413 Arabidopsis genomic gene sequences, complete and partial, were identified.

Each rice genomic gene sequence (see above) was used as a query in computer-assisted sequence similarity searches using the programs BLAST (National Center for Biotechnology Information, Bethesda, MD; accessed during August 1994) (21) and FASTA (UWGCG) (20). Sequence similarity >30 bp between the noncoding regions of two or more rice genes were compiled, aligned, and displayed using the UWGCG programs, LINEUP, PILEUP, and BOXSHADE, respectively. Sequence similarity identified on the basis of simple sequence repeats were not considered in this report.

DNA Isolation and Amplification. Rice germplasm (*O. sativa* cv. indica, accession no. IR25587-109-3-3-3-3) was obtained from G. Khush (International Rice Research Institute, Philippines). Wild rice germplasm and tissue were acquired

Abbreviations: SINEs, short interspersed nuclear elements; LINEs, long interspersed nuclear elements; CoA, coenzyme A; LTP, lipid transfer protein; HSP, heat shock protein; HMGR, 3-hydroxy-3-methylglutaryl CoA reductase; TIR, terminal inverted repeat; *Os*, *Oryza sativa*.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U61383 and U61071).

†To whom reprint requests should be addressed at the present address, Department of Biology, McGill University, 1205 Dr. Penfield, Avenue, Montreal PQ, Canada H3A 1B1. e-mail: Thomas_Bureau@maclan.mcgill.ca.

from G. Kochert (University of Georgia, Athens) and G. Second (International Rice Research Institute, Philippines). Genomic DNA was isolated as described previously (22).

Primers used in PCR protocols were selected with the aid of a MS-DOS program written by S. Marillonett (University of Georgia, Athens). Oligonucleotides (Bethesda Research Laboratories) were synthesized corresponding to sites flanking putative mobile elements in the rice genes encoding for 4-coumarate-coenzyme A (CoA) ligase (locus name OS4CL; nt 1494–1516, 5′-CCTCTCGTCGGTGCGGATGGTGC-3′; nt 2555–2532, 5′-CCCTGCCTCAGTCATTCCATACCC-3′), salt tolerance protein (OSSALT; nt 221–242, 5′-GAGGCTTCTTTGGCAGACGTGG-3′; nt 1073–1052, 5′-GTAGTTGTGTAGGGCAATGGGG-3′), 16.9-kDa heat shock protein (RICHSEA; nt 552–571, 5′-CATGACCCAAGACCGAACCG-3′; nt 1503–1481, 5′-CTCCACACTCCCAAGTGCTGGTC-3′), homeobox protein (RICOSH1; nt 479–502, 5′-CAAGGCTCCACTCCTCTACTACCC-3′; nt 1101–1080, 5′-CTCCGACGACCCCGAAGTGGTG-3′), and 3-hydroxyl-3-methylglutaryl CoA reductase (RIC3H3M; nt 2751–2770, 5′-AGTTGCTGGCTGCTGTCTTC-3′; nt 4858–4835, 5′-GAAGCCTACCAAATCTGCTGGACC-3′). PCR was performed as described previously (7) using annealing temperatures ranging from 50 to 72°C. Amplified products were cloned into a TA vector (Invitrogen). Plasmid preparations were performed using a Qiagen Miniprep kit and sequenced at the Molecular Genetics Instrumentation Facility (University of Georgia, Athens) using an ABI 373 (Applied Biosystems) automated DNA sequencer.

RESULTS

A Computer-Based Survey Reveals Several Common Elements in Rice Genes. A total of 105 rice and 413 Arabidopsis partial and complete genomic gene sequences are currently in the GenBank and EMBL nucleic acid databases. These sequences were used as individual queries in computer-assisted sequence similarity searches of all entries in the two databases. The rice survey reveals 32 common sequences in noncoding regions of rice genes, whereas no common sequences were identified within the noncoding regions of Arabidopsis genes.

Several of the rice common sequences are members of the previously described *Tourist*, *Stowaway*, and p-SINE*1* mobile element families (Table 1). In addition to those previously reported, five new *Tourist* and *Stowaway* elements were identified in this survey. Furthermore, an element in the 5′-flanking region of the rice lipid transfer protein (LTP) gene is similar to an unnamed element previously reported in the first intron of the rice α-amylase 2A gene (25, 26). We refer to these two elements as the Amy/LTP family.

The other common sequences define five newly described interspersed element families designated *Gaijin* (Japanese for foreigner or tourist), *Castaway*, *Ditto*, *Wanderer*, and *Explorer* (Table 2). Members of some of these new element families were detected in rice expressed sequence tags, rice sequenced tagged sites, and non-rice genes. Interestingly, five rice gene sequences harbor two element insertions [α-amylase 2A, proliferating cellular nuclear antigen, salt tolerance (*SalT*), thioredoxin h, 16.9-kDa heat shock protein (HSP)], one has three elements [3-hydroxy-3-methylglutaryl CoA reductase (HMGR)], and another has four element insertions [starch granule-bound UDPglucose-glucosyl transferase (*Waxy*)] (Tables 1 and 2).

Evidence for a Mobile History. Demonstration of past mobile activity for some members of the five newly described element families was assessed using a PCR-based protocol designed to survey the element insertion sites in the genomes of wild relatives of *O. sativa*. PCR primers were synthesized corresponding to regions flanking an element identified in our computer-based survey. In the most optimal situations, these primers were anchored within conserved regions of exons. This approach allowed us to identify insertion polymorphisms that confirmed or determined the position of element termini and putative target site sequences.

Table 1. Previously identified transposon families associated with wild-type rice genes.

Element family name	Gene description	Ref.
Tourist	Phytochrome 18, responsive-to-abscisic acid (Rab) 16B, α-amylase 2A, phenylalanine ammonia lyase (PAL), oryzacystatin II, starch branching enzyme (*Sbe*) 3-hydroxy-3-methylglutaryl CoA reductase (HMGR)*	7, 9, 23
Stowaway	Heat shock protein (HSP) 82A, α-amylase C, α-amylase 3A, poliferating cellular nuclear antigen [(PCNA), 2 elements], amylose synthase [(*Waxy*), 2 elements], salt tolerance (*SalT*)*, prolamin (RP3)*†	10, 24
p-SINE*1*	*Waxy*, 2 elements	13
Amy/LTP‡	α-Amylase 2A, lipid transfer protein (LTP)*	25, 26

*Genes containing elements identified in this study.
†Sequence has not been published.
‡No previous name given.

The *Gaijin* Element Family. The significant sequence similarity shared between *Gaijin* family members made it possible to delimit the termini of each element by multiple sequence alignments (Fig. 1*A*). In addition, these alignments revealed the presence of terminal inverted repeats (TIRs) and 3-bp direct repeats (5′-TNN-3′) flanking each *Gaijin* element.

Gaijin-Os1 is located in the first intron of the *O. sativa* 4-coumarate-CoA ligase gene (27). The same region amplified from a different accession of *O. sativa* and a wild relative *O. longistaminata*, however, lacks *Gaijin-Os1*; whereas the element is present in the same region of *O. rufipogon* (Fig. 1*B*). The insertion polymorphisms not only precisely correspond to the predicted position of the element, the putative target site sequence is present only once in the sequences lacking *Gaijin-Os1*.

Gaijin is not limited to rice since one member, *Gaijin-So1*, was found in the 3′-untranslated region of a sugarcane glucose transporter cDNA. The previously sequenced cDNAs from two sugarcane glucose transporter gene family members (SGT1 and SGT2) share >99% sequence similarity throughout their coding regions (30). In contrast, the 3′-untranslated regions of these cDNAs share virtually no significant sequence similarity. This striking difference is due to the insertion of *Gaijin-So1* in SGT2 (Fig. 1*C*). *Gaijin-So1* accounts for 93% of the SGT2 3′-untranslated region and is truncated by a 16-bp poly(A) tail. Since the two cDNAs are virtually identical, the insertion of *Gaijin-So1* probably represents a very recent event.

The *Castaway* Element Family. Like *Gaijin*, *Castaway* family members share both sequence and structural similarity. Multiple sequence alignments reveal that each *Castaway* member sequence has a TIR sequence and a 3-bp target site with an apparent bias for the sequence 5′-TAA-3′ (Fig. 2*A*). *Castaway* also includes non-rice members as *Castaway-Zm1* was identified in the third intron of a maize actin gene (Table 2, Fig. 2*A*).

The 5′-flanking region of the *O. sativa SalT* gene harbors two putative mobile elements designated *Castaway-Os1* and *Stowaway-Os9* (10, 23). Only nine base pairs separate the two elements. A comparison of the *O. sativa* sequence and the PCR amplified homologous region of *O. eichingeri* reveals an insertion polymorphism that corresponds precisely with *Castaway-Os1* and *Stowaway-Os9* (Fig. 2*B*). As with the *Gaijin* example,

Table 2. Newly identified common elements associated with wild-type rice genes.

Element family name	Element* designation	Gene description†	GenBank locus name	Position (nt)‡	Ref.
Gaijin	*-Os1*	4-Coumarate-CoA ligase	OS4CL	in1 (1888-2027)	27
	-Os2	Serine carboxypeptidase (*Cbp3*)	RICCBP3	5′ (1125-1251)	28
	-Os3	Aspartic protease	RICAP	5′ (302-448)	§
	-Os4	EST	RICC107591	99-235	29
	-So1	Glucose transporter	SCFGLUTRAB	3′ UTR (1637-1726)	30
Castaway	*-Os1*	Salt tolerance protein (*SalT*)	OSSALT	5′ (332-695)	24
	-Os2	Thioredoxin h	RICRTH	5′ (121-476)	31
	-Zm1	Actin (*MAc1*)	MZEACT1G	in3 (1738-2096)	32
Ditto	*-Os1*	16.9 kDa HSP	RICHSEA	5′ (635-878)	33
	-Os2	Homeobox (*OSH1*)	RICOSH1	5′ (591-819)	34
Wanderer	*-Os1*	HMGR	RIC3H3M	in1 (4476-4704)	23
	-Os2	HMGR	RIC3H3M	5′ (696-925)	23
	-Os3	Prepro-glutelin	RICGLUTE	5′ (1146-1366)	35
	-Os4	α-Amylase	OSALAM	5′ (1838-2042)	36
	-Os5	H3 Histone pseudogene	OSHIS3PS	5′ (2-187)	37
	-Ol1	RFLP	U34601	(410-642)	§
Explorer	*-Os1*	16.9 kDa HSP	RICHSEA	5′ (999-1178)	33
	-Os2	Thioredoxin h	RICRTH	5′ (514-689)	31
	-Os3	STS	RICG1103A	66-225	38

*Abbreviations of host plants are as follows: *Os, Oryza sativa; So, Saccharum officinalis; Zm, Zea mays* spp. *mays; Ol, Oryza longistaminata.*
†Abbreviations of sequence or gene names are as follows: EST, expressed sequence tag; HSP, heat shock protein; HMGR, 3-hydroxy-3-methylglutaryl CoA reductase; RFLP, restriction fragment length polymorphism; STS, sequence tagged site.
‡Location of some elements were in introns (in), 5′ flanking regions (5′) or untranslated regions (UTR).
§Sequence has not been published.

the polymorphism allowed us to confirm the predicted ends of the *Castaway* and *Stowaway* elements as well as their putative target site sequences, 5′-TAA-3′ and 5′-TA-3′, respectively.

The *Ditto* Element Family. A pairwise alignment of the genomic regions harboring the two *Ditto* elements was not sufficient to delimit their termini. However, the termini of *Ditto-Os1* could be identified following the characterization of an insertion polymorphism that distinguishes the 5′-flanking regions of the *O. sativa* and *O. punctata* 16.9-kDa HSP genes (Fig. 3*A*) (33). *Ditto-Os1* also has structural features consistent with mobile elements, i.e., a 3-bp target site sequence 5′-TAA-3′ and a 15-bp TIR (Fig. 3*B*). A pairwise alignment between the defined *Ditto-Os1* element sequence and the region of the *O. sativa OSH1* gene harboring *Ditto-Os2* reveals that *Ditto-Os2* has probably sustained a deletion encompassing 10 bp of the 3′ TIR and the predicted 3 bp target site sequence duplication (Fig. 3*B*).

The insertion of *Ditto-Os2* was the cause of an insertion polymorphism identified in a pairwise sequence comparison between the *OSH1* promoters of *O. sativa* and *O. punctata* (Fig. 3*B*) (34). The *O. punctata* promoter lacks *Ditto-Os2* but still retains the target site sequence duplication separated by a single guanine residue. Matsuoka *et al.* (34) have recently mapped the transcription start site of *OSH1*. Interestingly, the putative TATA box sequence (indicated by black lines in Fig. 3*B*) of the *O. sativa OSH1* gene appears to be supplied by *Ditto-Os2* and is absent in the *OSH1* promoter from *O. punctata*.

The *Wanderer* Element Family. Like *Ditto*, the termini of *Wanderer* elements could not be clearly defined via multiple sequence alignments. An insertion polymorphism identified between the first intron of the HMGR gene of *O. sativa* and *O. officinalis* corresponds to the insertion of *Wanderer-Os1* (Fig. 4*A*) (23). From this information, it was clear that *Wanderer-Os1* has an imperfect 10-bp TIR and a target site sequence of 5′-TAA-3′. Most of the *Wanderer* elements identified in this study share significant sequence similarity to the subterminal region of *Wanderer-Os1*, yet none appear to have convincing similarity to the putative *Wanderer-Os1* termini (Fig. 4*B*).

Wanderer-Ol1 was identified in a *O. longistaminata* genomic sequence (pTA8100) closely linked to the rice disease resistance gene Xa-21 (39). *Wanderer-Ol1* is absent in the orthologous region of *O. sativa*. This polymorphism did not aid in the determination of the *Wanderer-Ol1* termini, since the insertion event was accompanied by flanking rearrangements, i.e., short duplications (data not shown). Rearrangements may have also resulted in the apparent degeneracy of *Wanderer-Os2-5* termini.

The *Explorer* Element Family. *Explorer* element termini have no apparent substructure as determined by multiple sequence alignments (Fig. 5). Although an insertion polymorphism was identified between the 5′-flanking region of the *O. sativa* and *O. punctata* 16.9-kDa HSP genes, it was significantly larger than the region corresponding to *Explorer-Os1* (33). Only one other *Explorer* element, *Explorer-Os2*, was amenable to our PCR survey. However, three primer sets used in multiple attempts to amplify the region harboring *Explorer-Os2* from other Oryza species proved unsuccessful. The termini of the *Explorer* elements, therefore, remain undefined.

DISCUSSION

Systematic Identification of Transposons near Gene Sequences. We present a systematic approach to the identification of mobile elements near wild-type genes of *O. sativa* (domesticated rice) and *A. thaliana*. Complete genomic gene sequences were retrieved and used as queries in individual computer-based sequence similarity searches of all entries in the GenBank and EMBL databases. Two or more genes sharing sequence similarity in noncoding regions were compiled. Multiple sequence alignments were performed to delimit the termini of each element group. In cases where the element termini could not be determined, or to confirm the element ends, a PCR-based screen of element insertion points in related genes of other Oryza species was performed (see *Results*). Insertion polymorphisms identified via this PCR screen provided evidence for a mobile history. Each element family was examined for features indicative of mobile elements such as direct repeats, TIRs, coding capacity, and polyA regions. This systematic analysis should be applicable to all organisms for which sufficient gene sequences are available.

Our survey reveals nine putative mobile element families with members that are in close association with several wild-

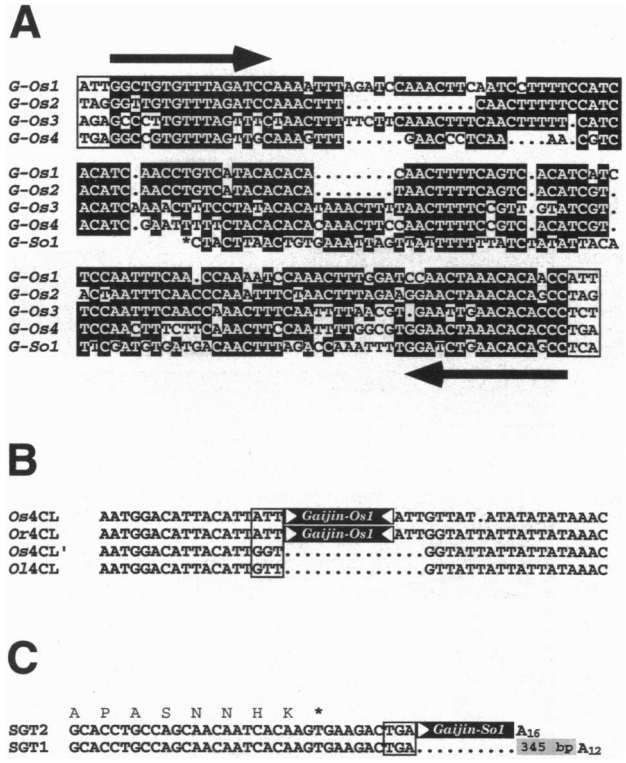

FIG. 1. (*A*) Multiple sequence alignments of *Gaijin* (*G*) family members. Solid arrows indicate the position of TIRs, and boxed regions indicate the putative target site of insertion. An asterisk indicates no further sequence available. Dotted lines are gaps. (*B*) Insertion polymorphisms between the first intron of the 4-coumarate-CoA ligase gene in two accessions of *O. sativa* (*Os*4CL and *Os*4CL′), *O. rufipogon* (*Or*4CL), and *O. longistaminata* (*Ol*4CL). Dotted lines refer to gaps. (*C*) Pairwise alignment of two sugarcane glucose transporter cDNAs (SGT1 and SGT2). Only the extreme 3′-region of each cDNA is shown. The translated sequence is indicated by non-boldface letters, and an asterisk indicates the position of the stop codon. Dotted lines are gaps, and the shaded box refers to a 345-bp region with no similarity to the SGT2 sequence.

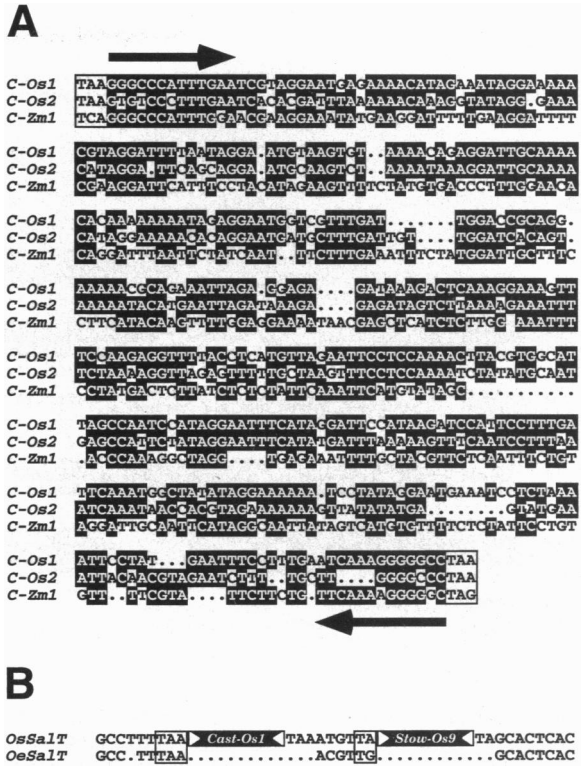

FIG. 2. (*A*) Multiple sequence alignments of *Castaway* (*C*) family members. Solid arrows indicate the position of TIRs, and boxed regions indicate the putative target site of insertion. Dotted lines are gaps. (*B*) Insertion polymorphism between the 5′-flanking region of the *SalT* gene in *O. sativa* (*OsSalT*) and *O. eichingeri* (*OeSalT*). This polymorphism is due to the insertion of two elements, *Castaway-Os1* (*Cast-Os1*) and *Stowaway-Os9* (*Stow-Os9*). The boxed region indicates the putative target site sequence. Dotted lines are gaps.

type rice genes. In total 37 elements were found using 105 rice gene sequences as queries. Five element families, designated *Gaijin*, *Castaway*, *Ditto*, *Wanderer*, and *Explorer* are first described in this study. The remaining four families, *Tourist*, *Stowaway*, Amy/LTP, and p-SINE*1*, were reported previously, although new members were identified in this study (Table 1). Insertion polymorphisms corresponding to members of four of the five new element families document a mobile history for these families (Figs. 1–4). The actual number of mobile elements associated with the rice queries is probably much higher since 105 gene sequences is merely a fraction of the number of genes in the rice genome and since our survey can only identify putative mobile elements that are shared by two or more gene sequences.

In a separate computer-based survey, we attempted to identify mobile elements near Arabidopsis gene sequences. Although there are approximately four times more Arabidopsis genomic gene sequences than rice genes in the GenBank and EMBL databases, no common elements were identified. Arabidopsis has the smallest known genome among higher plants (145 Mbp) and consequently has a very low amount of interspersed repetitive DNA (~4%) (40). Rice also has a small genome, the smallest among the cereal grasses (415–463 Mbp) (41). Given that transposable elements make up the majority of interspersed repetitive DNA, the results of our survey appears consistent with the small size of the Arabidopsis genome. The larger rice genome may reflect a relatively higher degree of mobile element activity during rice evolution resulting in the preponderance of elements associated with rice genes. Regardless, the results of our surveys highlight the difference in the composition of the noncoding regions of the genes of these two important plant systems.

Most Rice Elements in Genes Are Inverted-Repeat Elements and not SINEs or LINEs. SINEs and LINEs are abundant components of the noncoding regions of many mammalian nuclear genes (3, 4). SINEs are short nonautomonous mobile elements derived from structural RNA genes (3). Many SINEs have a polIII promoter, are terminated by a poly(A) tail, and have a random target site sequence of variable length (3, 4). LINEs, or non-long terminal repeat retrotransposons, are putative autonomous elements since they encode many of the genes required for mobility, i.e. reverse transcriptase (4). LINEs also are terminated by a poly(A) tail and, with some exceptions, do not have target site preference. There are no reports of SINEs or LINEs with TIRs.

With the exception of p-SINE*1*, the rice elements lack polIII promoters are not terminated by poly(A) tails nor do they have sequence similarity to previously sequenced SINEs or structural RNA genes. All of the rice elements lack coding capacity and have no sequence similarity to previously sequenced LINEs. Most rice elements have TIRs and either a 2- or 3-bp target site sequence. The vast majority of the rice elements identified in this study, therefore, are not SINEs or LINEs. Furthermore, computer-based amino acid-level searches using conceptual translations of plant LINEs did not reveal LINEs in rice or Arabidopsis nuclear genes.

The TIRs and/or target site sequence of some *Gaijin*, *Castaway*, *Ditto*, and *Wanderer* members are reminiscent of *Tourist*. Although not enough elements are available to derive

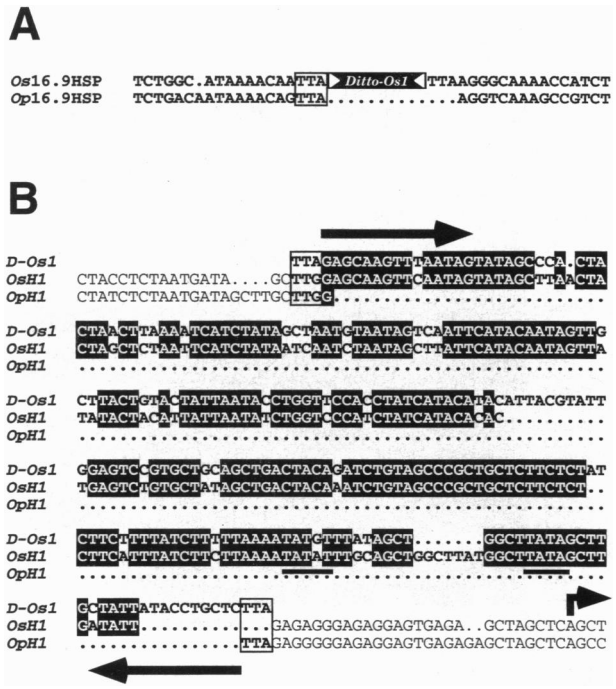

FIG. 3. (*A*) Insertion polymorphism between the 5′-flanking region of the 16.9-kDa HSP gene of *O. sativa* (*Os*16.9HSP) and *O. punctata* (*Op*16.9HSP). Boxed region indicates putative target site sequence of *Ditto-Os1*. Dotted lines are gaps. (*B*) Multiple sequence alignment between *Ditto-Os1* (*D-Os1*) and the 5′-flanking region of the maize *Knotted-1* homeobox gene homologue of *O. sativa* (*OsH1*) and *O. punctata* (*OpH1*). Boldface letters refer to *Ditto* related sequences. Non-boldface letters indicate sequences immediately flanking *Ditto* elements. Solid bent arrow indicates the position of the transcription start site of *OsH1*, and solid horizontal lines designate the putative TATA boxes. Solid arrows indicate the position of TIRs, and boxed regions indicate the putative target site of insertion. Dotted lines are gaps.

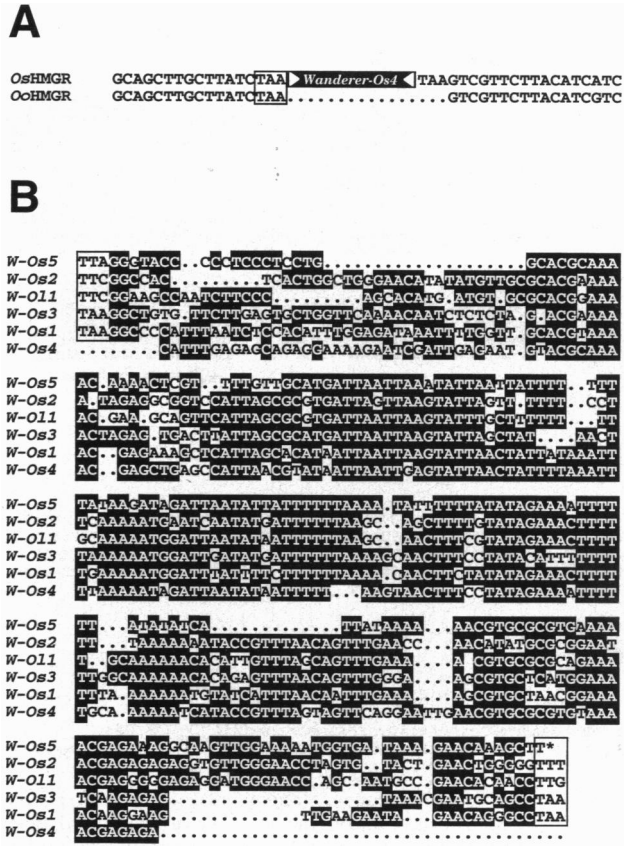

FIG. 4. (*A*) Insertion polymorphism between the first intron of the HMGR gene of *O. sativa* (*Os*HMGR) and *O. officinalis* (*Oo*HMGR). Boxed region indicates putative target site sequence of *Wanderer-Os1*. Dotted lines are gaps. (*B*) Multiple sequence alignment between *Wanderer* (*W*) elements. Boxed regions indicate the putative target site of insertion. The termini of *Wanderer-Os2*, *-Os4*, *-Os5*, and *-Ol1* could not be accurately determined and are shown as best estimates. An asterisk indicates that no further sequence is available, and dotted lines refer to gaps.

a statistically significant consensus TIR for any of the new rice element families, they do share a superficial resemblance to *Tourist* TIRs (5′-GGCCTTCTTCGGTT-3′) (Figs. 1–4) (7, 9). In fact, subtle variations in the *Tourist* TIR sequence have previously been used to define *Tourist* subfamilies (9). All of the new elements with identifiable TIRs have a 3-bp target site sequence and many share the preferred *Tourist* target site sequence 5′-TAA-3′. Both *Tourist* and *Stowaway* are characterized by the potential to form stable DNA secondary structure, i.e., a hairpin (7, 9, 10). None of the new rice elements, however, have this potential. Nevertheless *Gaijin*, *Castaway*, *Ditto*, and *Wanderer* are much more similar to *Tourist* than to any other previously identified mobile element family. For this reason, we refer to the group of elements comprising *Tourist*, *Stowaway*, *Gaijin*, *Castaway*, *Ditto*, and *Wanderer* as Miniature Inverted-repeat Transposable Elements or MITEs.

Potential Role of Rice Elements in Gene Evolution. Much of the speculation concerning the role of transposable elements in gene and genome evolution stems from the many examples of mutations caused by mobile element insertions. The phenotype of such mutants range from subtle changes in tissue specificity to dramatic alterations in the development and organization of tissues and organs (2). Mutant phenotypes can be generated from insertions not only in coding regions but also in noncoding regions such as 5′-flanking regions and introns (2, 43). However, there are only a few documented cases in which mobile elements have played a role in the evolution of normal or wild-type genes (5). Although the cis-factors involved in regulation of most sequenced rice genes have not been determined, the location of one rice MITE, *Ditto-Os2*, appears to correspond to the putative TATA box of the rice homologue of the maize homeobox gene *Knotted-1* (*Kn-1*) (Fig. 3*B*) (34). Since *Gaijin-So1* supplies almost the entire 3′-untranslated region of a sugarcane transporter cDNA, this element most likely supplies both the gene's polyadenylylation signal and site (Fig. 1*C*). In addition, *Tourist* and *Stowaway* elements may have contributed cis-factors to some plant genes (7, 9, 10). *Tourist-Zm11*, for instance, provides the core promoter for the maize auxin-binding protein gene (*Abp1*), which encodes the putative receptor for the plant hormone auxin (7, 44, unpublished data).

The extent to which mobile element insertions near genes is involved in the evolution of wild-type gene expression remains to be seen. The ability to resolve the presence of mobile element insertions near genes is partly hampered by the relatively small number of genomic sequences currently available and by the sensitivity of sequence similarity algorithms to align degenerate mobile element sequences. Nevertheless, the analysis of a limited number mammalian and rice genes clearly indicates that they have integral mobile element components (i.e., mammalian retroelements and rice MITEs) that may play a role in gene evolution. Furthermore, the proximity of mobile elements with very different mobile lifestyles to genes in mammals and rice may ultimately reflect a fundamental difference in animal and plant genome evolution.

We thank Lane Arthur and Shawn White for critical comments of our manuscript, Tom Holsten and Wenxue Zhai for technical assis-

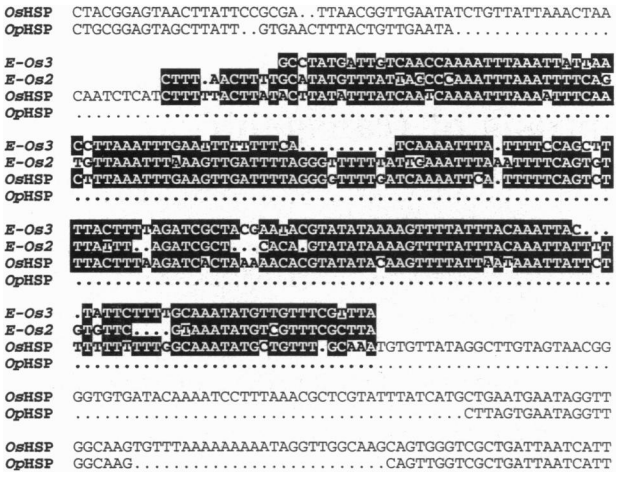

FIG. 5. Multiple sequence alignment between *Explorer-Os2* (*E-Os2*) and *-Os3* (*E-Os3*) and the 5′-flanking region of the 16.9-kDa HSP gene of *O. sativa* (*Os*HSP) and *O. punctata* (*Op*HSP). Boldface letters refer to *Explorer* related sequences. Non-boldface letters indicate sequences immediately flanking *Explorer-Os1* in *Os*HSP and the sequence of the homologous region of *Op*HSP. Dotted lines refer to gaps.

tance, and Michael Weise for advise on the use of the UWGCG computer programs. Financial support for this study was provided by grants from the National Institutes of Health (S.R.W. and P.C.R.) and the Rockefeller Foundation (S.R.W.).

1. Flavell, R. B. (1986) *Philos. Trans. R. Soc. Lond. Ser. B* **312,** 227–242.
2. McDonald, J. F. (1995) *Trends Ecol. Evol.* **10,** 123–126.
3. Eickbush, T. H. (1992) *New Biol.* **4,** 430–440.
4. Okada, N. (1991) *Trends Ecol. Evol.* **6,** 358–361.
5. McDonald, J. F. (1993) *Curr. Op. Genet. Dev.* **3,** 855–864.
6. Finnegan, D. J. (1989) *Trends Genet.* **5,** 103–107.
7. Bureau, T. E. & Wessler, S. R. (1992) *Plant Cell* **4,** 1283–1294.
8. Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N. & Machida, Y. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6562–6566.
9. Bureau, T. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1411–1415.
10. Bureau, T. E. & Wessler, S. R. (1994) *Plant Cell* **6,** 907–916.
11. White, S. E., Habera, L. F. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11792–11796.
12. Schwartz-Sommer, Z., Leclercq, L., Gobel, E. & Saedler, H. (1987) *EMBO J.* **6,** 3873–3880.
13. Mochizuki, K., Umeda, M., Ohtsubo, H. & Ohtsubo, E. (1992) *Jpn. J. Genet.* **67,** 155–166.
14. Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R. & Kumar, A. (1992) *Nucleic Acids Res.* **20,** 3639–3644.
15. Flavell, A. J., Smith, D. B. & Kumar, A. (1992) *Mol. Gen. Genet.* **231,** 233–242.
16. Voytas, D. F., Cummings, M. P., Konieczny, A., Ausubel, F. M. & Rodermel, S. R. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7124–7128.
17. Hirochika, H. & Hirochika, R. (1993) *Jpn. J. Genet.* **68,** 35–46.
18. Leeton, P. R. J. & Smyth, D. R. (1993) *Mol. Gen. Genet.* **237,** 97–104.
19. Deragon, J.-M., Landry, B. S., Pelissier, T., Tutois, S., Tourmente, S. & Picard, G. (1994) *J. Mol. Evol.* **39,** 378–386.
20. Devereaux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
22. Dellaporta, S. L., Wood, J. & Hicks, J. B. (1983) *Plant Mol. Biol. Rep.* **1,** 19–22.
23. Nelson, A. J., Doerner, P. W., Zhu, Q. & Lamb, C. J. (1994) *Plant Mol. Biol.* **25,** 401–412.
24. Claes, B., Dekeyser, R., Villarroel, R., Van den Bulke, M., Bauw, Van Montagu, M. & Caplan, A. (1990) *Plant Cell* **2,** 19–27.
25. Huang, N., Reinl, S. J. & Rodriguez, R. L. (1992) *Gene* **111,** 223–228.
26. Vignols, F., Lund, G., Pammi, S., Trémousaygue, D., Grellet, F., Kader, J.-C., Puigdomènech, P. & Delseny, M. (1994) *Gene* **142,** 265–270.
27. Zhao, Y., Kung, S. D. & Dube, S. K. (1990) *Nucleic Acids Res.* **18,** 6144.
28. Washio, K. & Ishikawa, K. (1992) *Plant Mol. Biol.* **19,** 631–640.
29. Sasaki, T., Song, J. Y., Kogaban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Muruyamakayano, E., Takiguchi, T., Takasuga, A., Niki, T., Ishimaru, K., Ikeda, H., Yamamoto, Y., Mukai, Y., Ohta, I., Miyadera, N., Havukkala, I. & Minobe, Y. (1994) *Plant J.* **6,** 615–624.
30. Bugos, R. C. & Thom, M. (1993) *Plant Physiol.* **103,** 1469–1470.
31. Ishiwatari, Y., Honda, C., Kawashima, I., Nakamura, S.-I., Hirano, H., Mori, S. Fujiwara, T., Hayashi, H. & Chino, M. (1995) *Planta* **195,** 456–463.
32. Shah, D. M., Hightower, R. C. & Meagher, R. B. (1983) *J. Mol. Appl. Genet.* **2,** 111–126.
33. Tzeng, S.-S., Yeh, K.-W., Chen, Y.-M. & Lin, C.-Y. (1992) *Plant Physiol.* **99,** 1723–1725.
34. Matsuoka, M., Ichikawa, H., Saito, A., Tada, Y., Fujimura, T. & Kano-Murakami, Y. (1993) *Plant Cell* **5,** 1039–1048.
35. Abe, K., Emori, Y., Kawasaki, H., Kondo, H., Suzuki, K. & Arai, S. (1989) *Agric. Biol. Chem.* **53,** 2969–2973.
36. Huang, N., Sutcliff, T. D., Litts, J. C., & Rodriguez, R. L. (1990) *Plant Mol. Biol.* **14,** 655–668.
37. Wu, S. C., Vegh, Z., Wang, X. W., Tan, C. C. & Dudits, D. (1989) *Nucleic Acids Res.* **17,** 3297.
38. Inoue, T., Zhong, H. S., Miyao, A., Ashikawa. I., Monna, L., Fukuoka, S., Miyadera, N., Nagamura, Y., Kurata, N., Sasaki, T. & Minobe, Y. (1994) *Theoret. Appl. Genet.* **89,** 728–734.
39. Ronald, P. C., Albano, B., Tabien, R., Abenes, L., Wu, K. S., McCouch, S. & Tanksley, S. D. (1992) *Mol. Gen. Genet.* **236,** 113–120.
40. Meyerowitz, E. M. (1994) in *Arabidopsis*, eds. Meyerowitz, E. M. & Somerville, C. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 21–36.
41. Arumuganathan, K. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 208–218.
42. Martinez, C. P., Arumuganathan, K., Kikuchi, H. & Earle, E. D. (1994) *Jpn. J. Genet.* **69,** 513–523.
43. Alleman, M. & Kermicle, J. L. (1993) *Genetics* **135,** 189–203.
44. Schwob, E., Choi, S.-Y., Simmons, C., Migliaccio, F., Ilag, L., Hesse, T., Palme, K. & Soll, D. (1993) *Plant J.* **4,** 423–432.