The Plant Cell, Vol. 6, 907–916, June 1994 © 1994 American Society of Plant Physiologists

Stowaway: A New Family of Inverted Repeat Elements Associated with the Genes of Both Monocotyledonous and Dicotyledonous Plants

Thomas E. Bureau [1] and Susan R. Wessler

Departments of Botany and Genetics, University of Georgia, Athens, Georgia 30602

Members of a new inverted repeat element family, named *Stowaway*, have been found in close association with more than 40 monocotyledonous and dicotyledonous plant genes listed in the GenBank and EMBL nucleic acid data bases. *Stowaway* elements are characterized by a conserved terminal inverted repeat, small size, target site specificity (TA), and potential to form stable DNA secondary structures. Some elements are located at the extreme 3′ ends of sequenced cDNAs and supply polyadenylation signals to their host genes. Other elements are in the 5′ upstream regions of several genes and appear to contain previously identified *cis*-acting regulatory domains. Although the *Stowaway* elements share many structural features with the recently discovered *Tourist* elements, the two families share no significant sequence similarity. Together, the *Stowaway* and *Tourist* families serve to define an important new class of short inverted repeat elements found in possibly all flowering plant genomes.

INTRODUCTION

The majority of interspersed repetitive DNA in eukaryotes has been suggested to be transposable elements or their remnants (Flavell, 1986). Moreover, some highly repetitive families of transposable elements are frequently associated with genes. Several human gene sequences, for instance, harbor the retroposon *Alu* (~10^6 copies per haploid genome) and the long interspersed nuclear sequence (LINE) L1 (~10^5 copies per haploid genome) (Berg and Howe, 1989). The proximity of transposable elements may influence the expression of the neighboring cellular genes by activating cryptic or supplying *cis*-acting regulatory regions (Clemens, 1987; Paulson et al., 1987; Baumruker et al., 1988; Stavenhagen and Robins, 1988; Banville and Boie, 1989; Chang-Yeh et al., 1991; Goodchild et al., 1992; Maichele et al., 1993).

Flowering plants have genomes that are on average much larger than those of other higher eukaryotes and are thought to have a correspondingly larger number of transposable elements (Bennett and Smith, 1991). Some known plant retrotransposons occur at high copy number in their host genomes (Grandbastien, 1992). The *del2* (dispersed element of lilies) element, for example, constitutes 4% of the lily genome (Leeton and Smyth, 1993). Previously, we have described a recent insertion of a mobile element, *Tourist-Zm1* (*Zea mays*), in a maize *waxy* allele (Bureau and Wessler, 1992). This element was found to be a member of another highly repetitive transposable element family associated with more than 30 wild-type genes of cereal grasses listed in nucleotide data bases (Bureau and Wessler, 1992, 1994). *Tourist* is characterized by terminal inverted repeats (TIRs), small size, target site preference (TAA), and potential to form stable DNA secondary structure. In this report, we describe a new family of transposable elements, named *Stowaway*, which are similar in structure but not in sequence to *Tourist* and are associated not only with listed gene sequences of cereal grasses but also with dicotyledonous plant genes. Furthermore, the fact that some *Stowaway* elements contain previously identified *cis*-acting regulatory regions provides evidence that this new family has contributed to the evolution of host gene expression.

[1] To whom correspondence should be addressed at the Department of Genetics, University of Georgia, Life Sciences Building, Athens, GA 30602.

RESULTS

Identification of *Stowaway* in Higher Plant Gene Sequences

The *Tourist-Sb5* (*Sorghum bicolor*) element, located at the extreme 5′ end of the sorghum phosphoenolpyruvate carboxylase CP21 gene sequence (Lepiniec et al., 1993; Bureau and Wessler, 1994), is interrupted by a 257-bp insertion (Figure 1). The presence of an imperfect TIR and a flanking 2-bp direct repeat (TA) suggests that this insertion, similar to *Tourist*, may be a transposable element. We have named this new element *Stowaway-Sb1*.

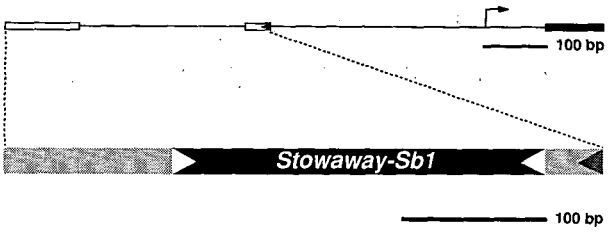

Figure 1. Position of *Stowaway-Sb1* within *Tourist-Sb1*.

The disrupted *Tourist-Sb1* (top diagram, open boxes) in the 5′ flanking region of the sorghum phosphoenolpyruvate carboxylase CP21 gene (transcription start site, bent arrow; 5′ coding sequence, black rectangle) has been expanded to show the position of *Stowaway-Sb1*. Triangles indicate the position of TIRs.

To determine if this element was a member of a larger family, computer-assisted sequence similarity searches of the GenBank (version 77.0) and EMBL (version 34.0) nucleic acid data bases were performed using *Stowaway-Sb1* as a query sequence. As new *Stowaway* elements were identified, these sequences were also used as queries. New *Stowaway* elements were defined as sequences that shared not only significant nucleotide similarity (>60% overall sequence similarity between any two elements), but also other structural features characteristic of the family, including TIR sequence similarity, target site duplication size, secondary structure, and overall length. Members of the *Stowaway* family that were identified in this way are listed in Table 1. Surprisingly, 47 plant sequences were identified as harboring *Stowaway* elements. Although several more degenerate sequences (<60% overall sequence similarity with another element and/or only partial structural similarity with *Stowaway*) were identified, only the best matches are presented. Whereas *Tourist* elements were found only within selected cereal grasses, the *Stowaway* family has a much wider distribution, with members in both monocotyledonous and dicotyledonous plants.

Although sequence similarity between elements ranges from 45 to 85% (Figures 2 and 3 and data not shown), *Stowaway* family members share several other features. First, *Stowaway* elements have a conserved 11-bp TIR with an overall consensus sequence of $C_{90}T_{88}C_{93}C_{79}C_{82}T_{92}C_{88}C_{69}G_{77}T_{89}T_{65}$ (numbers in subscript refer to the percent occurrence). Second, *Stowaway* elements are small, ranging from 80 to 323 bp. Whereas the monocotyledonous elements are variable in size, the dicotyledonous elements are considerably more homogenous (248 ± 24 bp). In general, the reported *Stowaway* elements found in monocotyledonous plant genes are more similar to one another than to elements found in dicotyledonous plant genes and vice versa (Figures 2 and 3 and data not shown). Third, *Stowaway* elements are AT rich (72 ± 5%). Fourth, *Stowaway* has a strong target site preference; ~85% of the *Stowaway* elements listed in Table 1 have TA targets (Figures 2 and 3 and data not shown). Among plant transposable elements characterized to date, only *Tourist* and *Stowaway* have target sequence preference (Bureau and Wessler, 1992, 1994). A TA target site sequence is also characteristic for members of the *IS630-Tc1* (transposon of *Caenorhabditis*) family of transposable elements (Doak et al., 1994). There is, however, no significant sequence similarity between *IS630-Tc1* family members and *Stowaway*. Fifth, *Stowaway* elements have a potential to form DNA secondary structures (Table 1; Figure 4). *Stowaway-Zm3*, for example, can be folded into a perfect hairpin except for a 1-bp mismatch. The FB elements of Drosophila (Smith and Corces, 1991) and Tc6 of *Caenorhabditis* (Dreyfus and Emmons, 1991) also have the potential to form hairpin-like structures but lack significant sequence similarity with *Stowaway*.

Evidence for Element Insertion

Although *Stowaway* family members have structural features of transposable elements, it was important to obtain evidence for element mobility in both monocotyledonous and dicotyledonous plants. To this end, two approaches were utilized. *Stowaway-St5* (*Solanum tuberosum*), *Stowaway-Ps1* (*Pisum sativum*), and *Stowaway-Le2* (*Lycopersicon esculentum*) have inserted into dicotyledonous plant genes that belong to gene families (paralogous loci) or are members of a group of homologous genes that have been previously isolated from closely related species (orthologous loci). Alignment of the sequences flanking these elements with paralogous and orthologous loci indicated that these elements correspond exactly with insertion polymorphisms (Figure 5). To provide evidence for element mobility in monocotyledonous plants, polymerase chain reaction (PCR) was employed to amplify introns harboring *Stowaway-Os3* (*Oryza sativa*) and *Stowaway-Zm3* from orthologous loci of closely related species or of cultivars within the same species. In each case, insertion polymorphisms were identified that corresponded precisely to the location of *Stowaway* and a TA target site duplication, thus revealing the site of a relatively recent insertion event (Figure 5).

Stowaway Elements Contain Previously Identified *cis*-Acting Regulatory Domains

A subset of the *Stowaway* elements that are located in the 5′ flanking regions of plant genes harbors previously identified *cis*-acting regulatory domains. Sequences in the *Stowaway-Le2* element (located within the promoter of a tomato gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase), for instance, contain ~50% of the sequences protected in a DNase footprinting assay (Manzara et al., 1993). In addition, *Stowaway-Le1* shares ~80% sequence similarity with *Stowaway-Le2* and occupies ~65% of a negative regulatory domain identified within the LAT59 promoter (−804 to −418, relative to the start of transcription) (Twell et al., 1991). Similarly, two putative embryogenesis-specific nuclear factors bind within the internal sequences of the *Stowaway-Dc2* (*Daucus carota*) element of carrot (Hatzopoulos et al., 1990), and

Table 1. *Stowaway* Elements Associated with Plant Gene Sequences Listed in the GenBank and EMBL Data Bases

Element[a]	Locus Name	Gene Description[b]	Position[c]	Size (bp)	Δ°G[d]	References[e]
Monocotyledonous Plants						
Os1	OSHSP82A	Heat shock protein 82A	in2 (2150)	150	−58.0	NA
Os2	RICAMYC	α-Amylase C	3′ (1667)	245	−70.1	Kim and Wu (1992)
Os3	OSWAXY	Starch synthase	in13 (3049)	234	−77.1	Wang et al. (1990)
Os4	OSWAXY	Starch synthase	3′ (4046)	204	−44.5	Wang et al. (1990)
Os5	OSPCNAGEN	PCNA	5′ (−432)	122	−47.1	Suzuka et al. (1991)
Os6	OSPCNAGEN	PCNA	5′ (−1315)	227	−91.4	Suzuka et al. (1991)
Os7	OSRAMY3A	α-Amylase 3A	in2 (405)	234	−60.7	Sutliff et al. (1991)
Zm1	ZM27KZNB	27-kD zein	3′UTR/3′ (830)	163	−24.7	Das et al. (1991)
Zm2	ZMAZ22KD	22-kD zein	3′ (4195)	156	−30.0	Thompson et al. (1992)
Zm3	ZMAYSPG	*P*	in2 (4760)	80	−54.5	Athma et al. (1992)
Zm4	M23537	10-kD zein	3′ (591)	153	−27.8	Kirihara et al. (1988)
Zm5	ZMGAPC4	GAP dehydrogenase	in3 (1444)	157	−38.4	Kersanach et al. (1994)
Sh1	SCFSCPEPCD	PEP carboxylase	in6 (2675)	267	−75.1	Albert et al. (1992)
Sb1	SVPEPCGX	PEP carboxylase	5′ (−469)	255	−50.7	Lepiniec et al. (1993)
Sb2	SVPEPCGX	PEP carboxylase	in6 (3231)	131	−42.3	Lepiniec et al. (1993)
Ta1	S117442	Metallothionein	5′ (−505)	167	−49.9	Kawashima et al. (1992)
Ta2	TAAAM254	α-Amylase 2/54	5′ (−458)	100	−40.5	Huttly et al. (1992)
Hv1	HVBKIN12G	Protein kinase	in2 (1747)	81	−30.7	Halford et al. (1992)
Hv2	BLYRCABG	Rubisco activase *RcaA*	in3 (839)	159	−40.3	Rundle and Zielinski (1991)
Hv3	BLYGLB2	1,3-1,4-β-Glucanase	in1 (166)	323	−87.8	Wolf (1991)
Hv4	BLYCLDAA	Cold-regulated protein	3′UTR (604)	49[f]	ND	Cattivelli and Bartels (1990)
Hv5a	HVPRP1A	Pathogenesis related (Hv-1a)	3′UTR (636)	123[f]	ND	Bryngelsson and Gréen (1989)
Hv5b	HVPRP1B	Pathogenesis related (Hv-1b)	3′UTR (647)	36[f]	ND	Bryngelsson and Gréen (1989)
Hv5c	HVPRP1C	Pathogenesis related (Hv-1c)	3′UTR (604)	93[f]	ND	Bryngelsson and Gréen (1989)
Dicotyledonous Plants						
Ps1	PEALCTN	Lectin	5′ (−1296)	275	−64.9	Mandaci and Dobres (1993)
Ps2	PEACAB80	*CAB* binding protein	5′ (−615)	276	−56.5	Timko et al. (1985)
Pc1	PCPR2G	Pathogenesis related (PR2)	5′ (−415)	243	−90.0	Van de Löcht et al. (1990)
Dc1	DCDC8	DC8	5′ (−514)	253	−65.9	Franz et al. (1989)
Dc2	S47635	DC59	5′ (−409)	249	−78.1	Hatzopoulos et al. (1990)
Bn1	BNEPSPG	EPSP synthase	in5 (1835)	220	−43.6	Gasser and Klee (1990)
Bn2	BNAC2PROMO	Cruciferin	5′ (−1147)	247	−41.7	Breen and Crouch (1992)
Sa1	SASCHSG	Chalcone synthase	5′ (−334)	260	−57.7	Batschauer et al. (1991)
St1	STWIN12G	Wound induced	5′ (−139)	259	−53.1	Stanford et al. (1989)
St2	STRBCS1	Rubisco small subunit	in1 (337)	285	−49.2	NA
St3	STPROINI	Proteinase inhibitor	3′ (1393)	240	−37.0	Lee and Park (1989)
St4	POTPATA	Patatin	in2 (922)	259	−67.3	Mignery et al. (1988)
St5	STPATP1	Patatin pseudogene	in5 (4378)	226	−34.8	Pikaard et al. (1986)
St6	STPOAC58	Actin	in3 (1426)	167	−41.6	Drouin and Dover (1990)
Le1	LELAT59	Pollen maturation specific	5′ (−656)	260	−51.6	Twell et al. (1991)
Le2	LERBSS1	Rubisco small subunit	5′ (−275)	251	−40.2	Manzara et al. (1993)
Le3	TOMATPACA	Ca^{2+}-ATPase	5′ (−114)	290	−65.5	Wimmers et al. (1992)
Le4	TOMPHEAMLY	PAL	5′ (−494)	248	−43.1	NA
Nt1	NTT85A	Auxin binding protein	in2 (622)	225	−50.7	NA
Nt2	NTCHN50	Endochitinase	5′ (−209)	244	−46.1	Fukuda et al. (1991)
Np1	TOBPMA1A	H^+-ATPase	in2 (443)	247	−40.2	Perez et al. (1992)
Ns1	NTNIA2	Nitrate reductase	in1 (1041)	249	−46.5	Vaucheret et al. (1989)
Nr1	NRTY8	tRNA-Tyr	5′ (−62)	75[f]	ND	Fuchs et al. (1992)
Ph1	PETEPSP	EPSP synthase	5′ (−1025)	243	−39.0	Benfey et al. (1990)
Ph2	PHCHSD	Chalcone synthase D	5′ (−115)	221	−39.3	Koes et al. (1989)
Ph3	PHCHSG	Chalcone synthase G	in1 (884)	257	−62.7	Koes et al. (1989)

[a] *Os*, *Oryza sativa* (rice); *Zm*, *Zea mays* (maize); *Sh*, *Saccharum hybrida* (sugarcane); *Sb*, *Sorghum bicolor* (sorghum); *Ta*, *Triticum aestivum* (wheat); *Hv*, *Hordeum vulgare* (barley); *Ps*, *Pisum sativum* (pea); *Pc*, *Petroselium crispum* (parsley); *Dc*, *Daucus carota* (carrot); *Bn*, *Brassica napus* (rape seed); *Sa*, *Sinapis alba* (mustard); *St*, *Solanum tuberosum* (potato); *Le*, *Lycopersicon esculentum* (tomato); *Nt*, *Nicotiana tabacum* (tobacco); *Np*, *N. plumbaginifolia*; *Ns*, *N. sylvestris*; *Nr*, *N. rustica*; *Ph*, *Petunia hybrida*. *Stowaway-Os3* has been previously referred to as *Tnr1* (transposable element in rice) (Umeda et al., 1991).

[b] PCNA, proliferating cell nuclear antigen; GAP, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; *CAB*, chorophyll *a/b* binding protein; ESPS, 5-enolpyruvylshikimate-3-phosphate; PAL, phenylalanine ammonium-lyase.

[c] 5′, 5′ flanking region; 3′, 3′ flanking region; in, intron sequence; UTR, untranslated region. Positions (given in parentheses) are relative to the translation start site. The position of *Stowaway-Nr1* is relative to the start of transcription.

[d] Kilocalorie per mole; ND, not determined.

[e] NA, not available.

[f] Only partial sequence available.

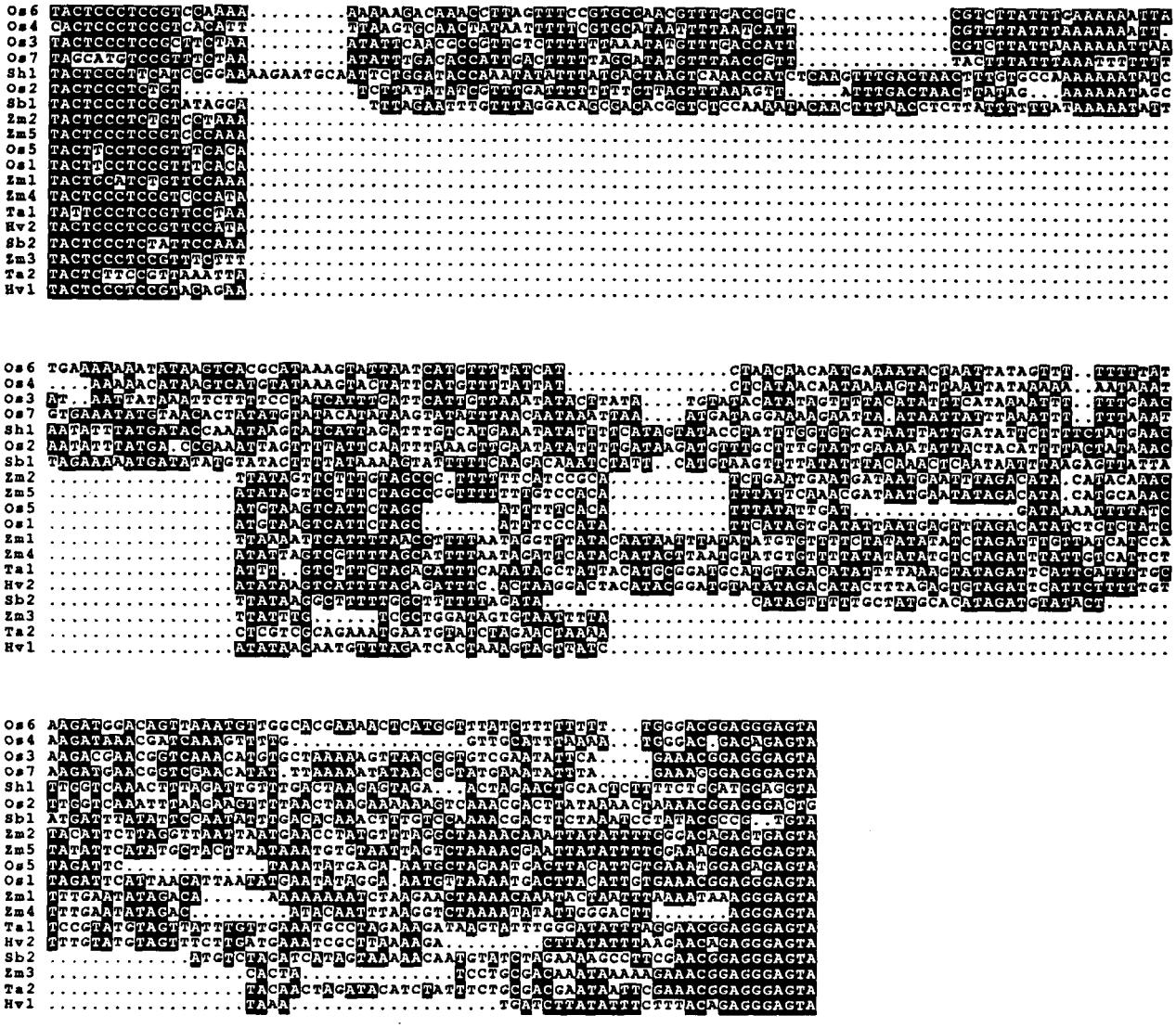

Figure 2. Multiple Sequence Alignments of *Stowaway* Elements Associated with Monocotyledonous Plant Genes.

Two base pairs immediately flanking the termini of each *Stowaway* element were included in the multiple alignment. The alignment is one of many possible optimal multiple sequence alignments and does not necessarily reflect the best pairwise relationship between any two *Stowaway* elements. *Stowaway-Hv3* and truncated elements (see Table 1) were omitted from the alignment. Conserved nucleotides are indicated by white letters on a black background.

Stowaway-St1 occupies ~45% of an upstream region important for wound inducibility in potato (Stanford et al., 1989, 1990).

In addition to the identification of *Stowaway* among genomic sequences, elements were also found in four barley stress-induced mRNAs (Table 1). Whereas one of these mRNAs was inducible by cold temperature stress (Cattivelli and Bartels, 1990), the remaining three transcripts (Hv-1a, Hv-1b, and Hv-1c) were derived from members of a gene family that encodes pathogenesis-induced thaumatin-like proteins (Bryngelsson and Gréen, 1989). The presence of *Stowaway* at the same position in all three Hv-1 transcripts indicates that insertion predates the amplification of this gene family. Interestingly, each Hv-1 transcript is polyadenylated at a different site within *Stowaway* sequences (Figure 6). This may indicate either that these related elements (>93% sequence similarity) have multiple sites for polyadenylation or that the minor sequence differences influence poly(A) site selection.

DISCUSSION

In this report, we describe an element family named *Stowaway*, which was first identified as an insertion in a member of another

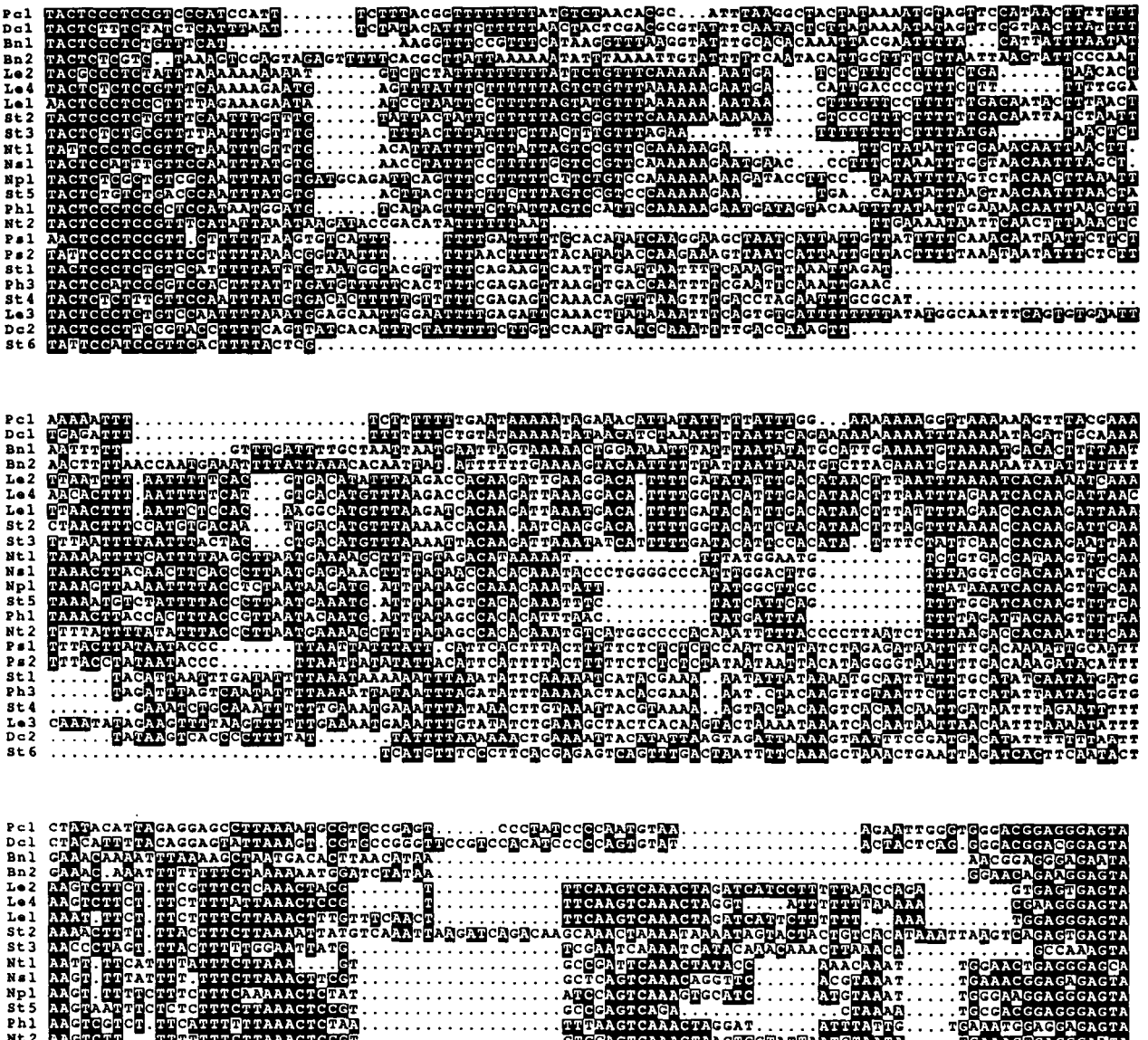

Figure 3. Multiple Sequence Alignment of *Stowaway* Elements Associated with Dicotyledonous Plant Genes.

Stowaway-Ph2, *Stowaway-Sa1*, and truncated elements (see Table 1) were omitted from the alignment. See legend to Figure 2. Abbreviations are as given in Table 1.

family of transposable elements, *Tourist*. More than 30 genes from several grasses harbor the transposable element *Tourist*; they are short (113 to 343 bp), have the potential to form DNA secondary structures, are AT rich, and have a preference for insertion at the trinucleotide TAA (Bureau and Wessler, 1992, 1994). *Tourist* elements have a mobile history because a maize mutant *waxy* allele was found to be caused by the recent insertion of *Tourist-Zm1*. In addition, the locations of *Tourist* elements in other genes correspond with insertion polymorphisms at orthologous loci. Despite the fact that *Stowaway* shares no significant sequence similarity to *Tourist*, these two families of elements have strikingly similar structural features. For instance, *Stowaway* elements are short, are AT rich, have the potential to form DNA secondary structures, and have target

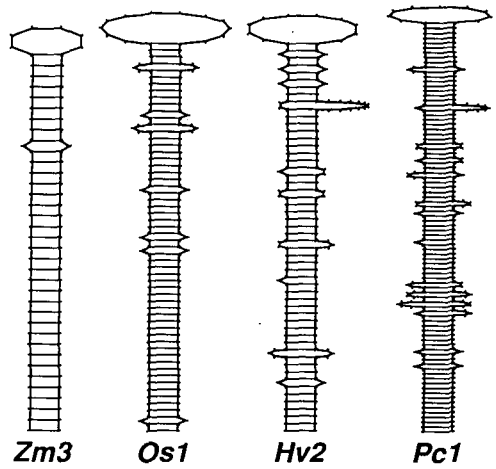

Figure 4. Stem-Loop Structures of *Stowaway-Zm3*, *-Os1*, *-Hv2*, and *-Pc1*.

The outline of each folded element is shown (horizontal lines correspond to a base pair). The dinucleotide direct repeats immediately flanking the elements were not included in the predicted DNA secondary structures. *Pc*, *Petroselinum crispum*.

insertion site preference. This suggests that the *Stowaway* and *Tourist* families are members of a larger class of elements that probably transpose by a similar mechanism. The presence of *Tourist* in maize, sorghum, rice, and barley indicates that *Tourist* is probably ubiquitous in the genomes of cereal grasses. Computer-assisted data base searches revealed that *Stowaway* elements are found in the genomes of both monocotyledonous and dicotyledonous plants, indicating that the *Tourist/Stowaway* superfamily of mobile elements is an important component of the genomes of possibly all flowering plants.

Mobility of some *Stowaway* elements is evident by their correspondence to insertion polymorphisms between orthologous and paralogous loci. Such polymorphisms also verify that the *Stowaway* target site is a dinucleotide (preferentially TA). Examples of *Stowaway* mobility given in Figure 5 indicate that element activity has occurred on an evolutionary time scale. For instance, *Stowaway-Os1* was identified in intron 2 of the heat shock protein 82A gene of domesticated rice, *O. sativa*, and other A genome–type rice species (Table 1 and data not shown). The absence of *Stowaway-Os1* in non-A-genome–type rice species (Figure 5D) suggests that the *Stowaway-Os1* insertion corresponds to the approximate divergence date of the A genome (~14 to 17 million years; Dally, 1988). In contrast, it appears that *Stowaway-Zm1* has transposed into intron 2 of the maize *P* gene much more recently because an insertion polymorphism was identified between orthologous loci of two maize inbred lines (Figure 5E).

The mechanism of mobility of the *Tourist/Stowaway* element superfamily remains unknown. The presence of TIRs is reminiscent of inverted repeat elements that transpose via a DNA intermediate or "cut-and-paste" mechanism (Bureau and

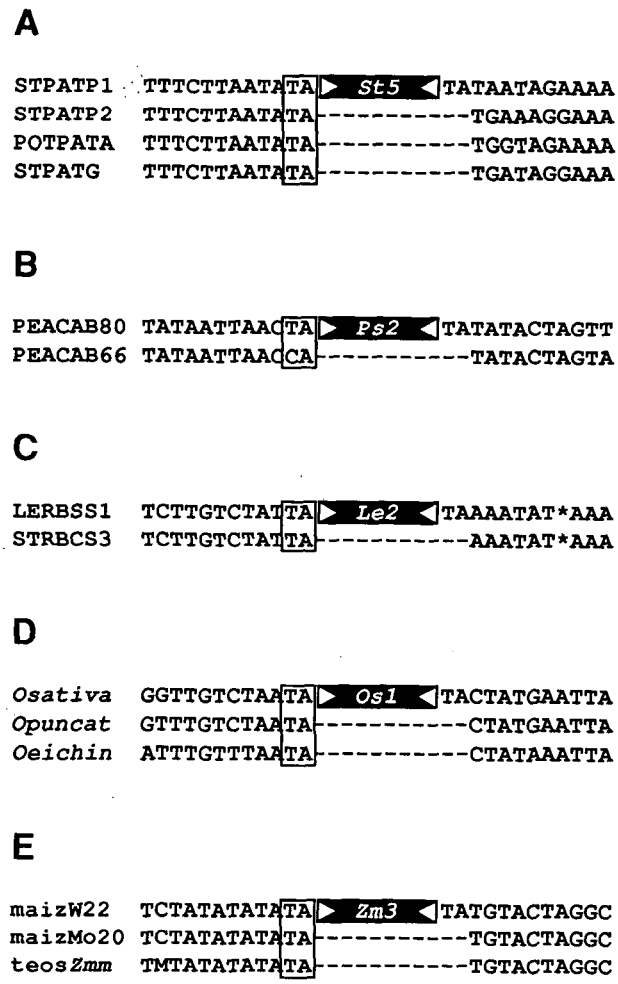

Figure 5. Polymorphisms Corresponding to *Stowaway* Insertions into Dicotyledonous and Monocotyledonous Genes.

(A) *Stowaway-St5* is located within intron 5 of the potato patatin pseudogene (STPATP1) but not in the corresponding position of three other members of the patatin gene family (STPATP2, POTPATA, and STPATG).
(B) *Stowaway-Ps2* is located in the 5′ flanking region of the pea *CAB80* (PEACAB80) gene but not in *CAB66* (PEACAB66), another member of the *CAB* gene family in pea.
(C) *Stowaway-Le2* is located in the 5′ flanking region of the tomato *rbcS1* gene (LERBSS1) but not in the corresponding position of the potato *rbcS3* gene (STRBCS3). An asterisk indicates a short variable region (LERBSS1, 7 bp; STRBCS3, 16 bp).
(D) *Stowaway-Os1* is located within intron 2 of the heat shock protein 82A gene of *Oryza sativa* (*Osativa*) but not in the wild rice species *O. puncata* (*Opuncat*) and *O. eichingeri* (*Oeichin*).
(E) *Stowaway-Zm3* is located within intron 2 of the maize *P* gene of the inbred line W22 (maizW22) but not in the Mo20 (maizMo20) inbred line or in the teosinte, *Z. mays* subsp *mexicana* (teos*Zmm*).
In all cases [**(A)** through **(E)**], significant sequence similarity extends past the region delimited. The presumed TA target sites are boxed, and dashed lines indicate gaps corresponding to the *Stowaway* element and one copy of the target site.

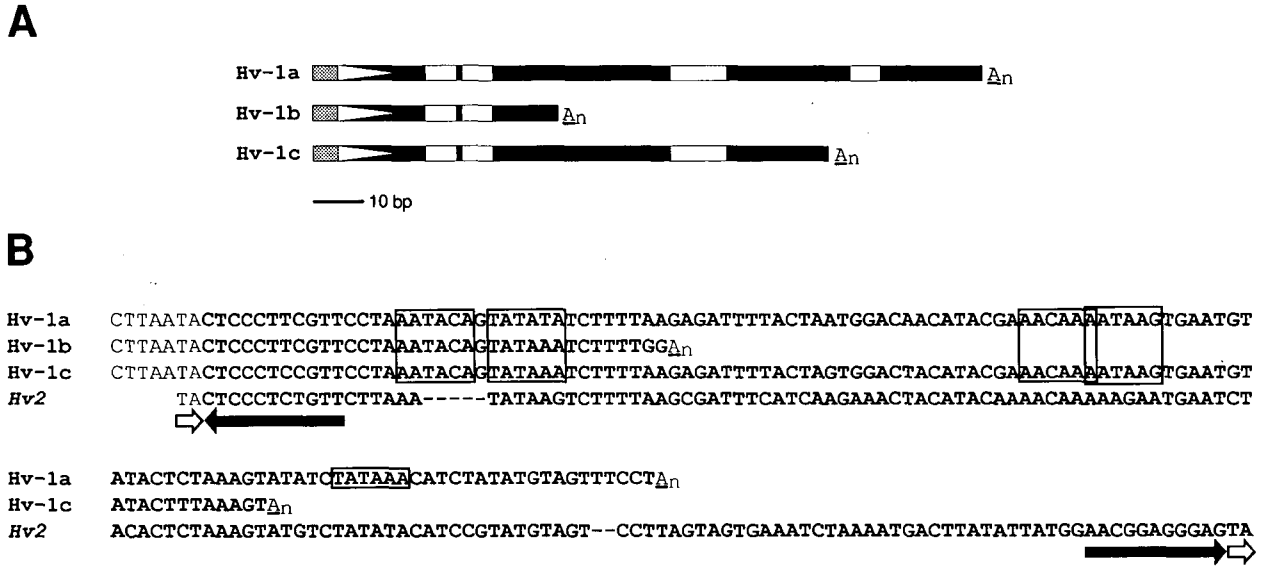

Figure 6. Polyadenylation Sites of Three Hv-1 Gene Family Members Occur in *Stowaway* Sequences.

(A) Schematic of the 3′ ends of three Hv-1 transcripts showing the positions of putative polyadenylation signals (open rectangles) in *Stowaway* sequences (black rectangles with open arrowheads extending to the poly[A] addition sites [A_n]).

(B) Sequences of the 3′ ends of each Hv-1 transcript aligned with the *Stowaway-Hv2* element (located within intron 3 of the barley Rubisco activase gene). The inverted repeat (black arrows), direct repeat (open arrows), putative polyadenylation signals (boxed sequences), and poly(A) tails (A_n) are indicated. Gaps (dashed lines) were introduced for optimal alignment.

Wessler, 1992, 1994). It cannot be ruled out, however, that *Tourist* and *Stowaway* are solo long terminal repeats (LTRs) because the LTRs of some retroelements also have TIRs, albeit short (~5 bp). Retroelements transpose via an RNA intermediate and do not excise. The absence of detectable excision events in our study does not necessarily support the notion that *Tourist* and *Stowaway* are solo LTRs because excision for these family of elements may be rare and/or precise.

Although *Stowaway* and *Tourist* elements share no significant sequence similarity to other previously reported transposable elements, some aspects of their structures are reminiscent of the *IS*630-Tc1 transposon superfamily (Berg and Howe, 1989; Dreyfus and Emmons, 1991; Doak et al., 1994). Members of this superfamily share significant transposase sequence similarity and, similar to *Stowaway*, have a TA target sequence preference. Interestingly, the 1.6-kb Tc6 element of *Caenorhabditis*, a Tc1-like element, consists of a 765-bp TIR and has the potential to form inverted repeat DNA secondary structures similar to that of *Tourist* and *Stowaway* elements (Dreyfus and Emmons, 1991). Because no *Stowaway* element identified to date contains a significant open reading frame that would encode a transposase, it would be premature to suggest that *Stowaway* is a member of the *IS*630-Tc1 superfamily. Furthermore, elements belonging to the *IS*630-Tc1 superfamily have been found in many diverse species but have not as yet been identified in plants.

The correspondence of *Stowaway* element location with previously identified *cis*-acting regulatory domains provides strong evidence that these elements have influenced the evolution of normal genes. For instance, *Stowaway-Hv4* and *Stowaway-Hv5* (*Hordeum vulgare*) provide polyadenylation signals and polyadenylation sites for their host genes. In addition, some elements (i.e., *Stowaway-Le1*, *Stowaway-Le2*, *Stowaway-Dc2*, and *Stowaway-St1*) may provide *cis*-acting regulatory regions to downstream genes. Previous reports of transposable elements supplying the *cis*-acting regulatory domains of normal genes are restricted to retrotransposons and retroposons. For example, the retrotransposons LTR-IS/MuRRS (murine retrovirus-related sequence), RTVL-H (retrovirus-like element with a histidine tRNA primer binding site), and THE-1 (transposon-like human element) provide polyadenylation signals for the mouse A1, human PLT (placental LTR terminated), and THE-1–containing genes, respectively (Paulson et al., 1987; Baumruker et al., 1988; Goodchild et al., 1992). The retroposon B2 has been identified in the 3′ ends of several mouse transcripts, and, in one example, provides alternative polyadenylation signals for the mouse γ-phosphorylase kinase gene (Clemens, 1987; Maichele et al., 1993). Furthermore, a VL30 (virus-like element encoding 30S RNA)–like retroviral insertion confers androgen sensitivity on the mouse sex-limited protein gene (Stavenhagen and Robins, 1988), and IAP (intracisternal-A particles)–derived solo LTRs supply the promoters of the rat oncomodulin and mouse MIPP (mouse IAP-promoted placental) genes (Banville and Boie, 1989; Chang-Yeh et al., 1991). Because the regulatory requirements of most plant genes that harbor *Stowaway* elements are poorly

characterized, we are uncertain of the extent of element involvement in the regulation of these genes.

Finally, it is important to note the role played by computer-assisted sequence similarity searches in the discovery of the *Stowaway* and *Tourist* families. Because many of these degenerate elements share less than 65% sequence identity over their length, standard hybridization protocols would not have been useful in identifying family members. In light of the enormous amount of sequences currently being generated by several genome projects, similar data base searches will undoubtedly lead to the identification of other elements and provide additional examples of the intimate association of mobile elements and normal genes.

METHODS

DNA Sequence Analysis

The UWGCG (University of Wisconsin, Madison, Genetics Computer Group) and IG (IntelliGenetics, Inc., Mountain View, CA) computer program suites were accessed through the BioSciences Computational Resource, University of Georgia, Athens, GA. Data base searches were conducted using the programs FASTDB (IG), FASTA (UWGCG), and BLAST (National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD) (Devereaux et al., 1984; Altschul et al., 1990). Pairwise and multiple alignments of element sequences were performed using the programs GAP and PILEUP (UWGCG), respectively. A gap penalty of 3.0 and gap length penalty of 0.3 were used, and complete elements were compared with their ends weighted. Minimum energy folding of element sequences was performed using the program FOLD (UWGCG) with DNA base pair and stacking energies as described previously (Breslauer et al., 1986). DNA secondary structures were visualized using the program SQUIGGLES (UWGCG).

DNA Manipulations

Oryza sativa (International Rice Research Institute [IRRI], Los Baños, The Philippines, accession number IR25587-109-3-3-3-3), *O. puncata* (IRRI accession number 103006), and *O. eichingeri* (IRRI accession number 101422) genomic DNA was obtained from G. Kochert (University of Georgia, Athens). *Zea mays* subsp *mexicana* (accession Iltis 28620) germplasm was acquired from J. Doebley (University of Minnesota, St. Paul, MN), and genomic DNA was isolated as previously described (Dellaporta et al., 1983). Oligonucleotides were synthesized corresponding to the sequences within or flanking intron 2 of the following genes. Nucleotide positions relative to the start of translation are given within parentheses. In the rice heat shock protein 82A gene (GenBank locus name OSHSP82A), the primer sequences are 5′-CATCTGGGGAGCAGCTTGGG-3′ (1558 to 1577) and 5′-TGAGGCGGCGCTCTTCAAGG-3′ (2481 to 2462); in the maize *P* gene (Athma et al., 1992), the primer sequences are 5′-ACACTGCGGACCGTGAGAGG-3′ (4510 to 4529) and 5′-GAGGTGGCTGGCGATCAGGG-3′ (5029–5010). Polymerase chain reaction (PCR) amplification was performed as previously described (Bureau and Wessler, 1992), except that an annealing temperature of 65°C was used. PCR products were checked by agarose gel electrophoresis and cloned into a TA plasmid vector (Invitrogen, San Diego, CA). Plasmid DNA isolation and dideoxy sequencing were performed using Qiagen (Chatsworth, CA) plasmid miniprep and Sequenase (United States Biochemicals) kits, respectively, as directed by the manufacturers.

ACKNOWLEDGMENTS

We thank Lane Arthur, Ronald Damiani, John McDonald, and Shawn White for critically reading this manuscript. In addition, we thank Michael Weise for computer advice, Andy Tull for greenhouse assistance, John Doebley for maize and teosinte germplasm, and Gary Kochert for rice genomic DNA. This study was funded by a grant from the National Institutes of Health (Grant No. GM32528) to S.R.W.

Received March 3, 1994; accepted April 13, 1994.

REFERENCES

Albert, H.A., Martin, T., and Sun, S.S.M. (1992). Structure and expression of a sugarcane gene encoding a housekeeping phosphoenolpyruvate carboxylase. Plant Mol. Biol. **20,** 663–671.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

Athma, P., Grotewold, E., and Peterson, T. (1992). Insertional mutagenesis of the maize *P* gene by intragenic transposition of *Ac*. Genetics **131,** 199–209.

Banville, D., and Boie, Y. (1989). Retroviral long terminal repeat is the promoter of the gene encoding the tumor-associated calcium-binding protein oncomodulin in the rat. J. Mol. Biol. **207,** 481–490.

Batschauer, A., Ehmann, B., and Schäfer, E. (1991). Cloning and characterization of a chalcone synthase gene from mustard and its light-dependent expression. Plant Mol. Biol. **16,** 175–185.

Baumruker, T., Gehe, C., and Horak, I. (1988). Insertion of a retrotransposon within the 3′ end of a mouse gene provides a new functional polyadenylation signal. Nucl. Acids Res. **16,** 7241–7251.

Benfey, P.N., Takatsuji, H., Ren, L., Shah, D.M., and Chua, N.-H. (1990). Sequence requirements of the 5-enolpyruvylshikimate-3-phosphate synthase 5′-upstream region for tissue-specific expression in flowers and seedlings. Plant Cell **2,** 849–856.

Bennett, M.D., and Smith, J.B. (1991). Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. Ser. B **334,** 309–345.

Berg, D.E., and Howe, M.M. (1989). Mobile DNA. (Washington, DC: American Society of Microbiology).

Breen, J.P., and Crouch, M.L. (1992). Molecular analysis of a cruciferin storage protein gene family of *Brassica napus*. Plant Mol. Biol. **19,** 1049–1055.

Breslauer, K.J., Frank, R., Blöcker, H., and Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA **83,** 3746–3750.

Bryngelsson, T., and Gréen, B. (1989). Characterization of a pathogenesis-related, thaumatin-like protein isolated from barley challenged with an incompatible race of mildew. Physiol. Mol. Plant Pathol. **35,** 45–52.

Bureau, T.E., and Wessler, S.R. (1992). *Tourist*: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell **4**, 1283–1294.

Bureau, T.E., and Wessler, S.R. (1994). 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.

Cattivelli, L., and Bartels, D. (1990). Molecular cloning and characterization of cold-regulated genes in barley. Plant Physiol. **93**, 1504–1510.

Chang-Yeh, A., Mold, D.E., and Huang, R.C.C. (1991). Identification of a novel murine IAP-promoted placenta-expressed gene. Nucl. Acids Res. **19**, 3667–3672.

Clemens, M.J. (1987). A potential role for RNA transcribed from B2 repeats in the regulation of mRNA stability. Cell **49**, 157–158.

Dally, A. (1988). Analyse cladistique de mutations de l'ADN chloroplastique et phylogénie des riz (section Eu-Oryza du genre Oryza). Collection Etudes et Thèses (dissertation) (Paris: Institut Français de Recherche Scientifique pour le Développement en Coopération [l'ORSTOM]).

Das, P.O., Ward, K., Ray, S., and Messing, J. (1991). Sequence variation between alleles reveals two types of copy correction at the 27-kDa zein locus of maize. Genomics **11**, 849–856.

Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983). A plant DNA minipreparation: Version II. Plant Mol. Biol. Rep. **1**, 19–22.

Devereaux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12**, 387–395.

Doak, T.G., Doerder, F.P., Jahn, C.L., and Herrick, G. (1994). A proposed superfamily of transposase genes: Transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA **91**, 942–946.

Dreyfus, D.H., and Emmons, S.W. (1991). A transposon-related palindromic repetitive sequence from *C. elegans*. Nucl. Acids Res. **19**, 1871–1877.

Drouin, G., and Dover, G.A. (1990). Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. J. Mol. Evol. **31**, 132–150.

Flavell, R.B. (1986). Repetitive DNA and chromosome evolution in plants. Philos. Trans. R. Soc. Lond. Ser. B **312**, 227–242.

Franz, G., Hatzopoulos, P., Jones, T.J., Krauss, M., and Sung, Z.R. (1989). Molecular and genetic analysis of an embryonic gene, DC 8, from *Daucus carota* L. Mol. Gen. Genet. **218**, 143–157.

Fuchs, T., Beier, D., and Beier, H. (1992). The $tRNA^{Tyr}$ multigene family of *Nicotiana rustica*: Genome organization, sequence analyses and expression *in vitro*. Plant Mol. Biol. **20**, 869–878.

Fukuda, Y., Ohme, M., and Shinshi, H. (1991). Gene structure and expression of a tobacco endochitinase gene in suspension-cultured tobacco cells. Plant Mol. Biol. **16**, 1–10.

Gasser, C.S., and Klee, H.J. (1990). A *Brassica napus* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase. Nucl. Acids Res. **18**, 2821.

Goodchild, N.L., Wilkinson, D.A., and Mager, D.L. (1992). A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. Gene **121**, 287–294.

Grandbastien, M.-A. (1992). Retroelements in higher plants. Trends Genet. **8**, 103–108.

Halford, N.G., Vicente-Carbajosa, J., Sabelli, P.A., Shewry, P.R., Hannappel, U., and Kreis, M. (1992). Molecular analyses of a barley multigene family homologous to the yeast protein kinase gene *SNF1*. Plant J. **2**, 791–797.

Hatzopoulos, P., Franz, G., Choy, L., and Sung, R.Z. (1990). Interaction of nuclear factors with upstream sequences of a lipid body membrane protein gene from carrot. Plant Cell **2**, 457–467.

Huttly, A.K., Phillips, A.L., and Tregear, J.W. (1992). Localisation of *cis* elements in the promoter of a wheat α-*Amy2* gene. Plant Mol. Biol. **19**, 903–911.

Kawashima, I., Kennedy, T.D., Chino, M., and Lane, B.G. (1992). Wheat E_c metallothionein genes. Like mammalian Zn^{2+} metallothionein genes, wheat Zn^{2+} metallothionein genes are conspicuously expressed during embryogenesis. Eur. J. Biochem. **209**, 971–976.

Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.-X., Martin, W., and Cerff, R. (1994). Five identical intron positions in ancient duplicated genes of eubacterial origin. Nature **367**, 387–389.

Kim, J.K., and Wu, R. (1992). Nucleotide sequence of a high-pI rice (*Oryza sativa*) amylase gene. Plant Mol. Biol. **18**, 399–402.

Kirihara, J.A., Petri, J.B., and Messing, J. (1988). Isolation and sequence of a gene encoding a methionine-rich 10-kDa zein protein from maize. Gene **71**, 359–370.

Koes, R.E., Spelt, C.E., van den Elzen, P.J.M., and Mol, J.N.M. (1989). Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. Gene **81**, 245–257.

Lee, J.S., and Park, J.S. (1989). Nucleotide sequence of a potato inhibitor I gene. Singmul Hakhoe Chi **32**, 69–78.

Leeton, P.R.J., and Smyth, D.R. (1993). An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol. Gen. Genet. **237**, 97–104.

Lepiniec, L., Keryer, E., Philippe, H., Gadal, P., and Crétin, C. (1993). Sorghum phosphoenolpyruvate carboxylase gene family: Structure, function and molecular evolution. Plant Mol. Biol. **21**, 487–502.

Maichele, A.J., Farwell, N.J., and Chamberlain, J.S. (1993). A B2 repeat insertion generates alternate structures of the mouse muscle γ-phosphorylase kinase gene. Genomics **16**, 139–149.

Mandaci, M., and Dobres, M.S. (1993). Sequence of a vegetative homolog of the pea seed lectin gene. Plant Physiol. **103**, 663–664.

Manzara, T., Carrasco, P., and Gruissem, W. (1993). Developmental and organ-specific changes in DNA-protein interactions in the tomato *rbcS1*, *rbcS2* and *rbcS3A* promoter regions. Plant Mol. Biol. **21**, 69–88.

Mignery, G.A., Pikaard, C.S., and Park, W.D. (1988). Molecular characterization of the patatin multigene family of potato. Gene **62**, 27–44.

Paulson, N.E., Matera, A.G., Deka, N., and Schmid, C.W. (1987). Transcription of a human transposon-like sequence is usually directed by other promoters. Nucl. Acids Res. **15**, 5199–5215.

Perez, C., Michelet, B., Ferrant, V., Bogaerts, P., and Boutry, M. (1992). Differential expression within a three-gene subfamily encoding a plasma membrane H^+-ATPase in *Nicotiana plumbaginifolia*. J. Biol. Chem. **267**, 1204–1211.

Pikaard, C.S., Mignery, G.A., Ma, D.P., Stark, V.J., and Park, W.D. (1986). Sequence of two apparent pseudogenes of the major potato tuber protein, patatin. Nucl. Acids Res. **14**, 5564–5566.

Rundle, S.J., and Zielinski, R.E. (1991). Organization and expression of two tandomly oriented genes encoding ribulose bisphosphate carboxylase/oxygenase activase in barley. J. Biol. Chem. **266**, 4677–4685.

Smith, P.A., and Corces, V.G. (1991). *Drosophila* transposable elements: Mechanisms of mutagenesis and interactions with the host genome. Adv. Genet. **29,** 229–300.

Stanford, A., Bevan, M., and Northcote, D. (1989). Differential expression within a family of novel wound-induced genes in potato. Mol. Gen. Genet. **215,** 200–208.

Stanford, A.C., Northcote, D.H., and Bevan, M.W. (1990). Spatial and temporal patterns of transcription of a wound-inducible gene in potato. EMBO J. **9,** 593–603.

Stavenhagen, J.B., and Robins, D.M. (1988). An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell **55,** 247–254.

Sutliff, T.D., Huang, N., Litts, J.C., and Rodriguez, R. (1991). Characterization of an α-amylase multigene cluster in rice. Plant Mol. Biol. **16,** 579–591.

Suzuka, I., Hata, S., Matsuoka, M., Kosugi, S., and Hashimoto, J. (1991). Highly conserved structure of proliferating cell nuclear antigen (DNA polymerase sigma auxiliary protein) gene in plants. Eur. J. Biochem. **195,** 571–575.

Thompson, G.A., Siemieniak, D.R., Sieu, L.C., Slightom, J.L., and Larkins, B.A. (1992). Sequence analysis of linked maize 22 kDa α-zein genes. Plant Mol. Biol. **18,** 827–833.

Timko, M.P., Kausch, A.P., Hand, J.M., Cashmore, A.R., Herrera-Estrella, L., Van den Broeck, G., and Van Montagu, M. (1985). Structure and expression of nuclear genes encoding polypeptides of the photosynthetic apparatus. In Molecular Biology of the Photosynthetic Apparatus, K.E. Steinback, S. Bonitz, C.J. Arntzen, and L. Bogorad, eds (New York, NY: Cold Spring Harbor Laboratory), pp. 381–396.

Twell, D., Yamaguchi, J., Wing, R.A., Ushiba, J., and McCormick, S. (1991). Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. Genes Dev. **5,** 496–507.

Umeda, M., Ohtsubo, H., and Ohtsubo, E. (1991). Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. Jpn. J. Genet. **66,** 569–586.

Van de Löcht, U., Meier, I., Hahlbrock, K., and Somssich, I.E. (1990). A 125-bp promoter fragment is sufficient for strong elicitor-mediated gene activity in parsley. EMBO J. **9,** 2945–2950.

Vaucheret, H., Kronenberger, J., Rouzé, P., and Caboche, M. (1989). Complete nucleotide sequence of the two homologous tobacco nitrate reductase genes. Plant Mol. Biol. **12,** 597–600.

Wang, Z.Y., Wu, Z.L., Xing, Y.Y., Zheng, F.G., Guo, X.L., Zhang, W.G., and Hong, M.M. (1990). Nucleotide sequence of the rice waxy *gene. Nucl. Acids Res.* **18,** 5898.

Wimmers, L.E., Ewing, N.N., and Bennett, A.B. (1992). Higher plant Ca^{2+}-ATPase: Primary structure and regulation of mRNA abundance by salt. Proc. Natl. Acad. Sci. USA **89,** 9205–9209.

Wolf, N. (1991). Complete nucleotide sequence of a *Hordeum vulgare* gene encoding (1→3,1→4)-β-glucanase isoenzyme II. Plant Physiol. **96,** 1382–1384.