Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 6562–6566, July 1993
Genetics

Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs

YASUSHI YOSHIOKA*, SHOGO MATSUMOTO†, SHOKO KOJIMA*, KAZUHIKO OHSHIMA‡, NORIHIRO OKADA‡, AND YASUNORI MACHIDA*§

*Department of Biology, School of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan; †Institute for Cell Biology and Genetics, Faculty of Agriculture, Iwate University, Ueda 3-18-8, Morioka 020, Japan; and ‡Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 227, Japan

Communicated by Roy J. Britten, March 18, 1993 (received for review January 5, 1993)

ABSTRACT **We have characterized a family of tRNA-derived short interspersed repetitive elements (SINEs) in the tobacco genome. Members of this family of SINEs, designated TS, have a composite structure and include a region structurally similar to a rabbit $tRNA^{Lys}$, a tRNA-unrelated region, and a TTG repeat of variable length at the 3′ end. Southern blot hybridization, together with a search of the GenBank data base, showed that various plants belonging to the families Solanaceae and Convolvulaceae contain sequences homologous to the TS family in the introns and flanking regions of many genes, whereas *Arabidopsis* in the family Cruciferae and several species of monocotyledonous plants do not. The TS family is widely involved in structural and genetic variations in the genomes of many plants that belong to the order Tubiflorae. All of nine sequences identified in a data base search are truncated at their 5′ regions and lack the tRNA-related region of the TS family. We characterized the entire sequence of the members of the TS family and found that this family can be categorized as a member of a group of SINEs with a $tRNA^{Lys}$-like structure, as can several animal SINEs. The TS family can be divided into two major subfamilies by analysis of diagnostic positions, and one of the subfamilies is clearly younger than the other. Amplification of many copies of the full sequence of the younger subfamily occurred during the recent evolution of the tobacco lineage. We also discuss mechanisms that could be involved in the generation of SINEs in animals and also in plants.**

A retroposon is a unit of information, initially present as in an RNA transcript, that has been incorporated back into the genome via a cDNA intermediate (1, 2). A mechanism that creates retroposons, called retroposition, was recently characterized as one that generates structural variability in eukaryotic genomes (1, 2). Retroposons can be divided into viral and nonviral superfamilies (2). Nonviral retroposons are classified into three main groups: processed retropseudogenes, long interspersed repetitive elements (LINEs), and short interspersed repetitive elements (SINEs) (2). Among the retroposons, SINEs are predominant and most mammalian genomes typically contain large numbers of SINEs (3). Recent findings confirm the notion that amplifications of SINEs continuously generate genetic and structural variations in the host genome even at the present time (2, 4, 5).

In 1985, three groups demonstrated that most mammalian SINEs are derived from tRNAs or their genes (1, 6–8). Except for human *Alu* and related families, most SINEs have a composite structure that consists of a region homologous to a tRNA, a tRNA-unrelated region, and an A- or (A+T)-rich region (8, 9). A major group of SINEs in the animal kingdom is considered to be derived from a $tRNA^{Lys}$, or to a tRNA species that is structurally related to $tRNA^{Lys}$, because of the marked structural resemblance of these SINEs to this tRNA (9). Because of the absence of such retroposons in avian and *Drosophila* genomes, it was believed initially that SINEs are restricted to mammalian species (1, 2). Recent reports from several laboratories have indicated, however, that the $tRNA^{Lys}$-related SINEs are widespread in the animal kingdom in vertebrates and invertebrates (9). In plant species, only a single SINE family, named p-SINE*1*, has been characterized. It is found in the rice genome (10) but it has no significant similarity to a tRNA.

We report here an example of a plant SINE family that may have been derived from a tRNA. This SINE family, designated the TS family, is probably a member of the class of SINEs with a $tRNA^{Lys}$-like structure (11) that have been characterized in animals.

MATERIALS AND METHODS

For cloning of the TS family sequences and DNA sequencing, a tobacco genomic library of *Nicotiana tabacum* L. var. NK326 (purchased from Clontech) was screened using the ^{32}P-labeled *Eco*RI–*Dra* I DNA fragment (Fig. 1) that contains the repetitive sequence present around the target site for integration of transferred DNA (T-DNA) in clone D-2T (12, 13). Three independent positive clones were obtained and were designated Nt2, Nt3, and Nt4. A 0.7-kb *Sau*3AI fragment of Nt2, a 0.78-kb *Sau*3AI–*Eco*RI fragment of Nt3, and a 0.56-kb *Sau*3AI–*Eco*RI fragment of Nt4 were detected by Southern hybridization (14) using the ^{32}P-labeled *Eco*RI–*Dra* I DNA fragment as a probe. These fragments were subcloned in the pUC18 vector (15). DNA sequencing was carried out by the dideoxy chain-termination method (16).

RESULTS

Characterization of Repetitive Sequences in Tobacco and Other Plants. Our laboratory previously reported that a repetitive sequence is present around the target site for integration of T-DNA (a specific DNA region of the Ti plasmid that can be transferred from *Agrobacterium* to plant chromosomes) (12). Fig. 1 shows a restriction map of the genomic DNA fragment that contains the T-DNA integration site. The previous study showed that a repetitive sequence is present in the *Hin*dIII–*Eco*RI fragment (Fig. 1) (12). When we hybridized *Eco*RI digests of genomic DNA from leaves of nontransformed tobacco with the DNA fragments a, b, and c

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.

Abbreviations: SINE, short interspersed repetitive element; T-DNA, transferred DNA.

§To whom reprint requests should be addressed.

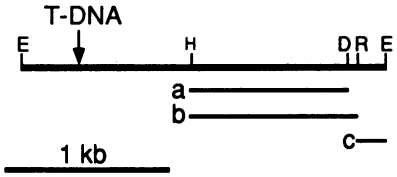

FIG. 1. Restriction map of the DNA fragments around the target site of T-DNA integration in the transformant D-2T (12). E, *Eco*RI; H, *Hin*dIII; D, *Dra* I; R, *Rsa* I. An arrow indicates the integration site of T-DNA.

as probes, fragments b and c hybridized to many heterogeneous DNAs, whereas fragment a hybridized to a single band of DNA (data not shown). This result shows that the *Dra* I–*Eco*RI fragment contains the repetitive sequence. The pattern of hybridization also indicates that this repetitive sequence is interspersed throughout the genome of tobacco.

We determined the nucleotide sequence of this fragment (Fig. 2, Nt1) and found that the sequence was flanked by direct repeats of TTTTATAACA but had none of the characteristic features of DNA transposons, such as terminal inverted repeats. A search of the GenBank data base (release 73.0) using the FASTA program (24) revealed that related sequences are present in introns and flanking regions of many genes in species belonging to the family Solanaceae, such as tobacco, tomato, and potato. Nine sequences related to that of Nt1 are shown in Fig. 2 (designated GNT to CHN). With the exception of GNT, all sequences are about 60 nucleotides shorter than that of Nt1, although three of them (STH2, HSF8, and CHN) have no obvious target site duplications, and ambiguities of one or two nucleotides are present around the 5′ and 3′ boundaries of these units.

Isolation of a tRNA-Derived Full-Length SINE from the Tobacco Genome. To isolate full-length members of this repeated family, we screened a tobacco genomic library using the *Dra* I–*Eco*RI fragment as a probe and isolated three additional members of this repetitive sequence from the tobacco genome (designated Nt2, Nt3, and Nt4). An alignment of these sequences showed that they constitute a typical SINE family, each of them composed of three distinct regions: a region homologous to a tRNA (thick line in Fig. 2) (see below), a tRNA-unrelated region, and a simple repeated unit of $(TTG)_n$ with a variable length of 7–45 nucleotides at the 3′ end (Fig. 2). The consensus sequence deduced from Nt1, Nt2, Nt3, Nt4, and GNT contains two blocks of sequences in the tRNA-related region that are homologous to the box A and box B sequences of the promoter for RNA polymerase III (23) (Fig. 2). We designated these SINEs as members of the TS (Tobacco SINE) family. As described below, *in vitro* transcription indicated that the G residue marked +1 (Fig. 2) is a site for initiation of transcription, and it may be the 5′ end of this repetitive sequence. Therefore, Nt2 and Nt3 contain the entire sequence that is typical of the

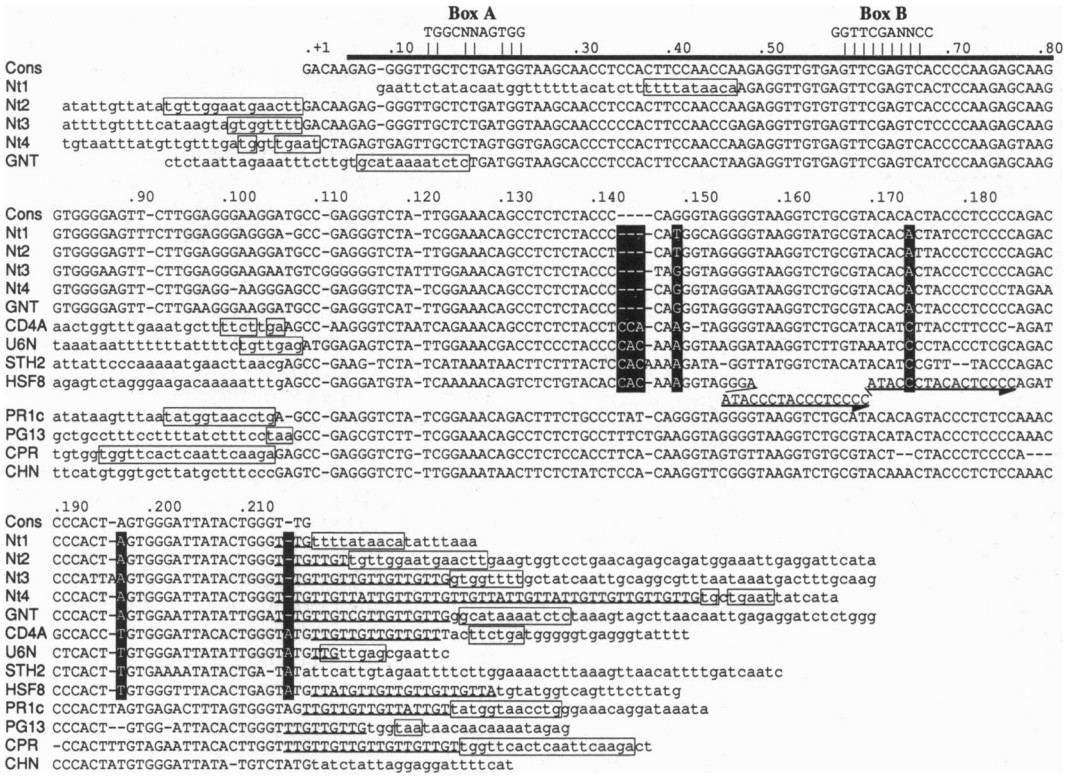

FIG. 2. Sequences of members of the TS family. Nt1 shows the nucleotide sequence of the repetitive sequence within the *Dra* I–*Eco*RI fragment shown in Fig. 1. Nt2, Nt3, and Nt4 show the nucleotide sequences of members of the TS family from clones Nt2, Nt3, and Nt4, respectively. GNT, CD4A, U6N, STH2, HSF8, PR1c, PG13, CPR, and CHN show the nucleotide sequences of members of the TS family associated with the genes indicated: GNT, the auxin-inducible gene GNT35 of tobacco (17); PR1c, the gene for pathogenesis-related 1c protein (PR1c) of tobacco (18); PG13, the G13 gene for the TGA1a-related protein in tobacco (19); CHN, the gene for class I chitinase CHN50 in tobacco (20); CPR, the gene for the chloroplast 29-kDa ribonucleoprotein in *Nicotiana sylvestris* (21); CD4A, the gene for ATP-dependent protease CD4A of tomato (22); HSF8, the gene for heat-stress transcription factor 8 in tomato (accession no. X67599); U6N, the gene for the U6 small nuclear RNA in potato (accession no. X60506); STH2, the STH-2 gene for the pathogenesis-related protein in potato (accession no. M29041). Cons shows a consensus nucleotide sequence deduced from the sequences of clones Nt1, Nt2, Nt3, Nt4, and GNT. Numbers above the consensus sequence represent nucleotide positions with respect to the first G residue, which was determined by the primer extension experiment, as shown in Fig. 4. Direct repeats are boxed. The tRNA-related region is indicated by a thick line. The TTG repeat is indicated by a thin line. Deletions are shown by bars. Box A and box B show consensus sequences for the RNA polymerase III promoter (23). Diagnostic positions are highlighted. Horizontal arrows indicate a tandem direct repeat.

TS family, whereas Nt4 lacks two nucleotides at the 5′ end. These sequences are flanked by perfect or nearly perfect short, direct repeats with lengths of 15 bp for Nt2, 8 bp for Nt3, and 7 bp for Nt4. The lengths of these units without their TTG repeats are 204, 207, and 202 nucleotides, respectively. The flanking regions of these repetitive sequences are slightly enriched for (A+T) pairs but differ completely from one another. Alignments of all of the sequences related to the TS family clearly showed that those detected by a search of GenBank are truncated forms and that the truncation sites are almost identical (positions 105–108, Fig. 2). It is possible that truncation may be a general feature of one subfamily of the TS sequences (see below).

tRNA-Related Sequence of the TS Family. In a computer-assisted similarity search, we found that of plant tRNA species reported to date, $tRNA^{Phe}$ (*Arabidopsis*) (25) and $tRNA^{Trp}$ (wheat) (26) are the most similar to the tRNA-related region of the TS family and they are 64% and 63% identical to this region, respectively (Fig. 3*A*). Since the sequences of several tRNAs in angiosperms ($tRNA^{Asp}$, $tRNA^{Glu}$, $tRNA^{Leu}$, $tRNA^{Lys}$, and $tRNA^{Thr}$) are not yet available, we cannot conclude that $tRNA^{Phe}$ and $tRNA^{Trp}$ are definitively the most similar to the TS SINE. The TS family also exhibits considerable sequence homology to a rabbit $tRNA^{Lys}$ (27). Fig. 3*B* shows the sequences and structural similarities between the tRNA-related region of members of the TS family and this $tRNA^{Lys}$ from rabbit (27). Of the 76 nucleotides in the tRNA-related region, 48 are identical to those in the tRNA (63% identity). It should also be noted that, in addition to box A and box B (Fig. 2), several nucleotides conserved and semiconserved in all tRNA molecules are present in the sequence of the TS family. For example, the conserved CU dinucleotide on the 5′ side of the anticodon loop of all tRNAs is retained at positions 38 and 39 in the TS family. The semiconserved AA dinucleotide on the 3′ side of the anticodon loop and the semiconserved GGU trinucleotide in the "extra" loop are also retained at positions 43 and 44 and 51–53 of the TS family, respectively.

A

```
tRNAPhe  5'-G-CGGGGAUAGCUCAGU-UGGGAGAGCG---UCAGACUGA-AG
           *  **** * **** *  *** * ***     **  ***    *
TS family 5'-G-AGGGGUU-GCUCUGA-UGGUA-AGCAACCUCC-ACUUCCA-
           . * *   **  ** *  * ****   **    **    *** ***
tRNATrp  5'-GGAUCCGUG-GCGCA-AUUGGUAGCGCGU-CUG--ACU-CCAG

         AUCUGA-AGGUCGCGUGUUCGA-UCCACGCUCACC-GCACCA-3'
         * *  * **** * * ****** * *** * **   ***
         ACCA-AGAGGUUGUGAGUUCGAGU-CACCC-CAAGAGCAAG-3'
         * ** * ****** * ****** * ***   *  *  **
         AUCAGA-AGGUUGCGUGUUCGA-UUCACGU-CGGGUCCACCA-3'
```

B

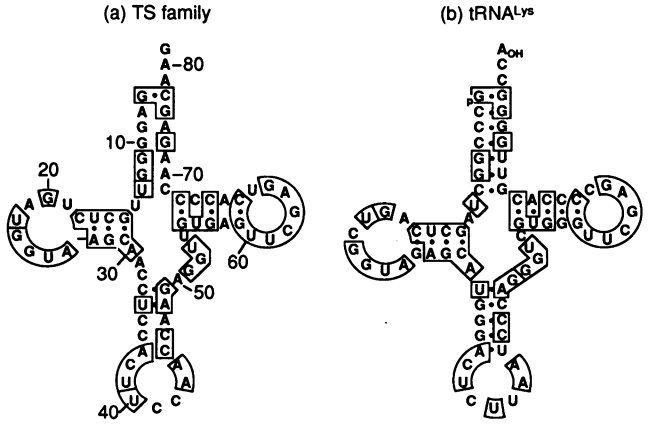

FIG. 3. Sequence and structural comparisons of the tRNA-related region of members of the TS family and tRNAs. (*A*) Sequence homologies between two plant tRNAs and the tRNA-related region of the TS family. The tRNA-related region of the consensus sequence of the TS family (Fig. 2) is aligned with *Arabidopsis* $tRNA^{Phe}$ (25) and wheat $tRNA^{Trp}$ (26). Identical sequences are indicated by asterisks. (*B*) Sequence and structural homologies between the tRNA-related region of members of the TS family (a) and rabbit $tRNA^{Lys}$ (b). Numbers along the sequence of the TS family are nucleotide sequence coordinates of the consensus sequence in Fig. 2. The sequence of $tRNA^{Lys}$ is taken from ref. 27. Identical sequences are boxed. Deletions are shown by bars. Only the four unmodified nucleotides are shown.

The TS Family Can Be Divided into Two Major Subfamilies. We found five diagnostic positions in the sequences of the TS family that allow the TS family to be clearly divided into two major subfamilies, designated the TSa and TSb subfamilies. The TSa subfamily includes Nt1, Nt2, Nt3, Nt4, and GNT, whereas the TSb subfamily includes CD4A, U6N, STH2, and HSF8 (Table 1). The nucleotides at the five diagnostic positions—namely, 143–145, 149, 174, 197, and 215—are specific to each subfamily, as highlighted in Fig. 2. It is possible that the TSa subfamily can be further subdivided into smaller subfamilies but this hypothesis cannot be validated at present because of the small size of the sample. The diagnostic nucleotides in PR1c, PG13, CPR, and CHN exhibit patterns intermediate between those of the TSa and TSb subfamilies. For example, in PR1c, the bases at positions 149, 174, and 197 reflect the pattern in TSa, whereas those at positions 143–145 and 215 reflect that in TSb. It also remains to be determined whether these clones belong to subfamilies intermediate between TSa and TSb or whether they reflect divergence of the TSb subfamily generated by the accumulation of neutral mutations in these sequences (28).

The consensus sequence was deduced separately for TSa and TSb, and their average sequence divergences were calculated to be 4.8% and 14.1%, respectively. These values indicate that the TSb subfamily is far older than the TSa subfamily.

Transcription of the TS Family *in Vitro*. We performed *in vitro* transcription in a HeLa cell extract. Fig. 4*A* shows an example of an assay in which the cloned Nt2 DNA (Fig. 2) was transcribed *in vitro* to produce a transcript of 350 nucleotides. In view of the sensitivity of the transcription to α-amanitin, it appears that this RNA is transcribed by RNA polymerase III. When total genomic DNA of tobacco was used as a template, it did not serve as a template for RNA polymerase III (data not shown), consistent with the frequent observation of 5′ truncated forms of the TS family. A primer extension experiment was performed to identify the site(s) of initiation of the transcription *in vitro*. As shown in Fig. 4*B*, three extension products were detected. The initiation site deduced from the longest extension product was assigned to the G residue at position +1, which is assumed to be the 5′ end of members of the TS family (Fig. 2).

Table 1. Nucleotides in the diagnostic positions

		Position*						
Class	Location	143	144	145	149	174	197	215
TSa	Nt1	—	—	—	T	A	A	—
	Nt2	—	—	—	T	A	A	—
	Nt3	—	—	—	G	A	A	—
	Nt4	—	—	—	G	A	A	—
	GNT	—	—	—	G	A	A	—
TSb	CD4A	C	C	A	A	C	T	A
	U6N	C	A	C	A	C	T	A
	STH2	C	A	C	A	C	T	A
	HSF8	C	A	C	A	C	T	A

*Numbers indicate the positions shown in Fig. 2.

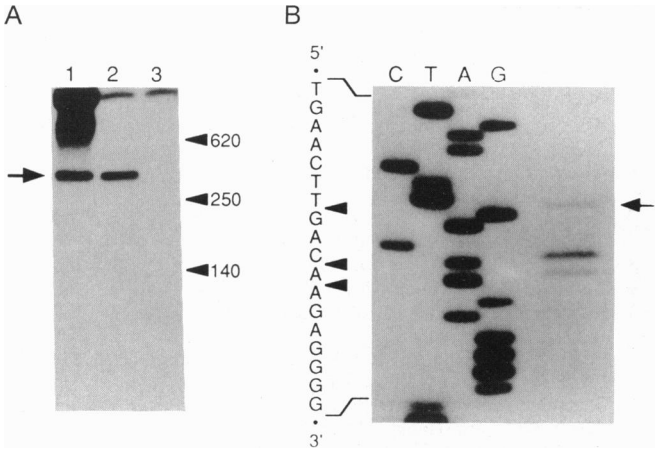

FIG. 4. Transcription *in vitro* of the TS family by a HeLa cell extract. (*A*) An extract of HeLa cells was prepared as described elsewhere (29) with slight modification (30). RNA was synthesized *in vitro* essentially as described elsewhere (29). Transcription of Nt2 DNA was performed in the absence (lane 1) or the presence of α-amanitin at 2 μg/ml (lane 2) or 200 μg/ml (lane 3). The transcript is indicated by the arrow and the positions of the size markers are indicated by numbers (nucleotides). (*B*) Determination of the site of initiation of transcription by primer extension analysis (right lane). The transcript generated *in vitro* was isolated from a denaturing 10% (wt/vol) polyacrylamide gel (31). The solution of RNA was mixed with 1 pmol of 5'-end-labeled synthetic oligonucleotide P1 (5'-GGTGACTCGAACACACAACC-3'; sequence complementary to nucleotides 51–70 in Nt2; see Fig. 2). The mixture was heated to 100°C for 2 min and then cooled on ice. RNA was synthesized with 33 units of reverse transcriptase under conditions recommended by the supplier (BRL). The transcript was subjected to electrophoresis on a 6% polyacrylamide/7 M urea gel with DNA sequencing markers. The longest extension product is indicated by an arrow.

Identification of Sequences Similar to the TS Family in the Solanaceae. Quantitative dot-blot hybridization showed that, assuming a tobacco cell contains 1.6×10^9 bp of DNA per haploid genome (32), at least 5.0×10^4 copies of the TS family are present in the haploid genome (data not shown). To examine the distribution of sequences similar to the TS family in the genomes of various plants, we performed Southern hybridization experiments. Sequences that can be hybridized with the TS family were detected in the genomes of plants belonging to the family Solanaceae and in the genome of *Pharbitis nil* (morning glory) but not in those of *Arabidopsis thaliana* and several monocotyledonous plants (data not shown). The systematic classification of plant species used in these experiments is shown in Table 2. These results indicate that many copies of sequences homologous to the TS family are interspersed throughout the genomes of species belonging to the order Tubiflorae.

DISCUSSION

Characteristics of the TS Family. In this study, we characterized a series of SINEs of tobacco, which were designated members of the TS family. The sequence of the TS family contains a region homologous to a few specific tRNAs and a tRNA-unrelated region (see Fig. 2). In addition, there is a TTG repeat of variable length at the 3' end. Most SINEs in the animal kingdom have A- or (A+T)-rich sequences at their 3' ends (1, 2, 14). The presence of the TTG repeat at the 3' end is a characteristic feature of the TS family. Rogers (1, 33) has proposed a model for the generation of the 3' tail of a SINE, in which the 3' tailing by telomerase of a nick in the chromosomal target for insertion of the SINE is postulated. Because the telomeric motifs from various organisms have been shown to be (G+T)-rich (34), the TTG repeat of the TS family may have been generated by telomerase at a certain step in its retrotransposition. Telomeric structures and telomerase from tobacco remain, however, to be characterized experimentally.

Diagnostic Positions Reveal Unique Features of the TS Subfamilies. Britten *et al.* (35) and Jurka and Smith (36) showed that the human *Alu* repeats that arose at earlier times shared correlated blocks of nucleotides that were different from the current consensus sequence at diagnostic positions. Using the same strategy, we identified two major subfamilies in the TS family. Unexpectedly, we found that all members of the TS family of full length (Nt2 and Nt3) or nearly full length (Nt4 and GNT) fall into the TSa subfamily, whereas several other members (CD4A, U6N, STH2, and HSF8) with truncated sequences fall into the TSb subfamily. The results suggest that members of the TSb subfamily tended to be truncated specifically at positions 105–108 during retroposition. From the average sequence divergences of the two subfamilies, the source gene of TSb seems to be far older than that of TSa. Indeed, the source gene of TSa may have been generated from TSb by accumulation of mutations at five diagnostic positions. Introduction of mutations at these positions may have altered the secondary structure or the sensitivity to nucleases of the transcript from TSb, thereby creating repeated units of TSa of full length during retroposition. These results also suggest that the first amplification of the TSa subfamily, which gave rise to dispersion of sequences of full length, occurred relatively recently in the evolution of the tobacco lineage.

Possible Mechanisms for the Generation of the TS Family in Plants. To our knowledge, a tRNA-derived SINE in plants has not been reported previously. It is fairly well established

Table 2. Plant species analyzed

Class	Order	Family	Genus	Species	Common name	Existence of the TS family
Dicotyledoneae	Tubiflorae	Solanaceae	*Nicotiana*	*tabacum*	Tobacco	+
			Nicotiana	*sylvestris*		+
			Solanum	*tuberosum*	Potato	+
			Lycopersicon	*esculentum*	Tomato	+
			Datura	*stramonium*		+
			Petunia	*hybrida*	Petunia	+
			Hyoscyamus	*niger*		+
		Convolvulaceae	*Pharbitis*	*nil*	Morning glory	+
	Rhoeadales	Cruciferae	*Arabidopsis*	*thaliana**		−
Monocotyledoneae	Glumiflorae	Gramineae	*Oryza*	*sativa**	Rice	−
			Triticum	*aestivum**	Wheat	−
			Zea	*mays**	Maize	−

*Unpublished data.

that most animal SINEs have evolved from tRNAs (for review, see ref. 9). It is remarkable that the tobacco TS family is composed of three distinct blocks as are animal SINEs, a feature that suggests that a general or similar mechanism may be involved in the initial generation of SINEs in animals and plants. Among animal SINEs, members of several families of SINEs are very similar to a vertebrate $tRNA^{Lys}$, being categorized as a superfamily of $tRNA^{Lys}$-related SINEs (11). Recently, it was found that five $tRNA^{Lys}$-related SINEs from distant species, such as rodent, fish, tortoise, and squid, have similar sequence blocks in their tRNA-unrelated regions and that the sequences complementary to these blocks are present in the U5 regions of several retroviruses, whose primer for reverse transcription is a $tRNA^{Lys}$. On the basis of these findings, a model for the generation of SINEs has been proposed, in which SINEs are derived from a "strong stop DNA" with a primer $tRNA^{Lys}$ that is an intermediate in the process of reverse transcription of retroviruses (11). If we assume that a similar mechanism is involved in the generation of the tobacco TS family, the most likely candidate for the origin of the tRNA-related region of the TS family is a primer tRNA of plant retrotransposons because retroviruses have yet to be demonstrated in plants.

The distribution of copia-like retrotransposons in higher plants has been studied extensively and they have been shown to be ubiquitous in such plants (37, 38). All plant retrotransposons characterized to date, including tobacco Tnt1, utilize $tRNA^{Met}$ as a primer for reverse transcription (39–41). Since the tRNA-related region of the TS family exhibits structural similarities to plant $tRNA^{Phe}$, $tRNA^{Trp}$, and animal $tRNA^{Lys}$ (Fig. 3), it is unlikely that the TS family originated from any of the plant retrotransposons reported to date. The present results might predict the possible presence of an unknown plant retrotransposon that can utilize a tRNA as a primer that might be an ancestor of the tRNA-related region of the TS family.

Another possible mechanism for the generation of the TS family involves horizontal transmission of a member of the superfamily of $tRNA^{Lys}$-related SINEs from an animal to a plant. With regard to horizontal transmission among many plant species, such a mechanism has already been proposed to be involved in the generation and distribution of the copia-like retrotransposons of plants (37, 38, 42). On the basis of the finding that plant cellular RNAs are encapsulated by the coat protein of tobacco mosaic virus (43), Hirochika and Hirochika (42) postulated the plant-virus-mediated transmission of an RNA transcript of a retrotransposon. They also proposed that insects play a role in transmission of retrotransposons because a number of plant viruses are known to be transmitted by insects. In addition to horizontal transmission among plant species, it has been suggested that a copia-like element was transmitted between *Arabidopsis* and *Drosophila* (44). The insect-mediated infection of plant viruses has also been postulated to account for the transmission of the copia-like element between this plant and the fruit fly (42). A similar mechanism may have contributed to horizontal transmission of SINEs between the animal and the plant kingdoms.

We are grateful to Dr. S. Iida (Science University of Tokyo, Japan) for the gift of the genomic DNA of *Pharbitis nil*. This research was supported in part by a Grant-in-Aid for General Scientific Research and a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture of Japan and by a grant for Pioneering Research Projects in Biotechnology from the Ministry of Agriculture, Forestry, and Fisheries of Japan.

1. Rogers, J. (1985) *Int. Rev. Cytol.* **93**, 187–279.
2. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
3. Singer, M. F. (1982) *Cell* **28**, 433–434.
4. Muratani, K., Hada, T., Yamamoto, Y., Kaneko, T., Shigeto, Y., Ohue, T., Furuyama, J. & Higashino, K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11315–11319.
5. Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W. & Collins, F. S. (1991) *Nature (London)* **353**, 864–866.
6. Lawrence, C. B., McDonnel, D. P. & Ramsey, W. J. (1985) *Nucleic Acids Res.* **13**, 4239–4252.
7. Daniels, G. R. & Deininger, P. L. (1985) *Nature (London)* **317**, 819–822.
8. Sakamoto, K. & Okada, N. (1985) *J. Mol. Evol.* **22**, 134–140.
9. Okada, N. (1991) *Trends Ecol. Evol.* **6**, 358–361.
10. Mochizuki, K., Umeda, M., Ohtsubo, H. & Ohtsubo, E. (1992) *Jpn. J. Genet.* **67**, 155–166.
11. Ohshima, K., Koishi, R., Matsuo, M. & Okada, N. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6260–6264.
12. Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y. & Machida, Y. (1990) *Mol. Gen. Genet.* **224**, 309–316.
13. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
14. Southern, E. M. (1985) *J. Mol. Biol.* **98**, 503–517.
15. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
17. Van der Zaal, E. J., Droog, F. N. J., Boot, C. J. M., Hensgens, L. A. M., Hoge, J. H. C., Schilperoort, R. A. & Libbenga, K. R. (1991) *Plant Mol. Biol.* **16**, 983–998.
18. Ohshima, M., Harada, N., Matsuoka, M. & Ohashi, Y. (1990) *Nucleic Acids Res.* **18**, 182.
19. Fromm, H., Katagiri, F. & Chua, N.-H. (1991) *Mol. Gen. Genet.* **229**, 181–188.
20. Van Buuren, M., Neuhaus, J. M., Shinshi, H., Ryals, J. & Meins, F. (1992) *Mol. Gen. Genet.* **232**, 460–469.
21. Ye, L., Li, Y., Fukami-Kobayashi, K., Go, M., Konishi, T., Watanabe, A. & Sugiura, M. (1991) *Nucleic Acids Res.* **19**, 6485–6490.
22. Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J. S., Dalrymple, B., Kuramitsu, H., Shiroza, T. & Foster, T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3513–3517.
23. Galli, G., Hofstetter, H. & Birnstiel, M. L. (1981) *Nature (London)* **294**, 626–631.
24. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
25. Akama, K. & Tanifuji, S. (1990) *Plant Mol. Biol.* **15**, 337–346.
26. Ghosh, K. & Ghosh, H. P. (1984) *Nucleic Acids Res.* **12**, 4997–5003.
27. Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., RajBhandary, U. L. & Gross, H. J. (1979) *Eur. J. Biochem.* **97**, 305–318.
28. Kimura, M. (1983) in *The Neutral Theory of Molecular Evolution*, eds. Osawa, S. & Honjo, T. (Cambridge Univ. Press, Cambridge), pp. 175–186.
29. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3856–3859.
30. Talkington, C. A., Nishioka, Y. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7132–7136.
31. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
32. Meyrowitz, E. M. & Pruitt, R. E. (1985) *Science* **229**, 1214–1218.
33. Rogers, J. (1983) *Nature (London)* **305**, 101–102.
34. Blackburn, E. H. (1991) *Nature (London)* **350**, 569–573.
35. Britten, R. J., Baron, W. F., Stont, D. B. & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4770–4774.
36. Jurka, J. & Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 475–478.
37. Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R. & Kumar, A. (1992) *Nucleic Acids Res.* **20**, 3639–3644.
38. Voytas, D. F., Cummings, M. P., Konieczny, A., Ausubel, F. M. & Rodermel, S. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7124–7128.
39. Voytas, D. F. & Ausubel, F. M. (1988) *Nature (London)* **336**, 242–244.
40. Grandbastien, M. A., Spielman, A. & Caboche, M. (1989) *Nature (London)* **337**, 376–380.
41. Hirochika, H., Fukuchi, A. & Kikuchi, F. (1992) *Mol. Gen. Genet.* **233**, 209–216.
42. Hirochika, H. & Hirochika, R. (1993) *Jpn. J. Genet.* **68**, 35–46.
43. Rochon, D. & Siegel, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1719–1723.
44. Konieczny, A., Voytas, D. F., Cummings, M. P. & Ausubel, F. M. (1991) *Genetics* **127**, 801–809.