Plant Molecular Biology (2006) 61:505–514
DOI 10.1007/s11103-006-0026-7

A SINE family widely distributed in the plant kingdom and its evolutionary history

Jeffrey A. Fawcett[1,*], Taihachi Kawahara[1], Hitoshi Watanabe[2] and Yasuo Yasui[1]
[1]*Laboratory of Crop Evolution, Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwa-kecho, Sakyoku, Kyoto 606-8502, Japan (*author for correspondence; e-mail jaf@kais.kyoto-u.ac.jp);* [2]*Kyoto Herbal Garden, Takeda Pharmaceutical Company Limited, Takenouchicho 11, Ichijoji, Sakyoku, Kyoto 606-8134, Japan*

Received 7 September 2005; accepted in revised form 10 February 2006

Key words: evolution, plant, retroelement, SINE

Abstract

The distribution and evolution of Au SINE in plants were examined. Au SINE is a short interspersed element first identified in *Aegilops umbellulata*, a close relative of wheat. The Au SINE was previously found in species such as wheat, maize, tobacco, and tomato, but not in rice. In this study, we first searched public databases, and next examined the presence of Au in a broad range of plant species by PCR using internal primers of Au. Although Au is likely to be absent from many species including rice, it was identified in many Gramineae, Solanaceae, and Fabaceae species, and also in a basal angiosperm species, *Asimina triloba*. Phylogenetic studies suggest that Au SINE originated before the divergence of monocots and eudicots. Au SINE sequences of *Asimina*, *Triticum*, *Zea*, *Nicotiana*, *Lotus*, *Medicago*, and *Glycine* were aligned and compared. Although sequences of Au were highly conserved among distantly related species, every Au element in *Glycine* had a 16 bp deletion and its 3′ end differed from sequences of other species. This type of Au could only be found in *G. max*, and not in other species including other Fabaceae species such as *M. truncatula* and *L. japonicus*. This is the first report of a plant SINE family present in multiple lineages, and the evolution of Au SINE in the plant kingdom, especially in Gramineae and Fabaceae is discussed.

Introduction

SINEs are retroelements that are thought to be present in most eukaryotic genomes. In spite of their potential to cause damage to their residing genome, it is thought that they have played their role in the evolutionary history of eukaryotes (Schmid, 1998; Weiner, 2002; Kazazian, 2004). Many SINE families have been identified in the animal kingdom, and recently two superfamilies of SINEs; CORE-SINE and V-SINE were found to be widespread in vertebrates (Gilbert and Labuda, 1999; Ogiwara *et al.*, 2002). In plants, only a few SINE families have been identified and their distribution is somewhat limited.

Yasui *et al.* (2001) identified a novel SINE family Au in *Aegilops umbellulata*, a close relative of wheat. The Au SINE is of about 180 bp, and like most other tRNA-derived SINEs, consists of a tRNA-related region including A and B boxes for internal RNA polymerase III promoters. This tRNA-related region is followed by a region unrelated to tRNA, but does not have any AT-rich region in its 3′ end. Au sequences end with a stretch of T residues, and are flanked by target site duplications, a characteristic feature of retroelements. Au was shown to be present in various monocotyledonous plants such as wheat, barley, maize, and also in dicotyledenous plants such as tobacco and tomato. However, Au could not be identified in rice, which is in the same

family Gramineae as wheat, barley, and maize (Yasui *et al.*, 2001).

This discontinuous distribution of Au prompted us to make a more detailed study of Au in order to understand its evolutionary history. We searched many plant species for the presence of Au by database search and PCR. We suspected that Au may have originated in the common ancestors of monocots and eudicots and therefore many gymnosperms and basal angiosperms were sampled. Since it was previously shown that Au is likely to be absent from rice, while present in other Gramineae species including maize, five species closely related to maize and rice were examined. Here we report the distribution of Au in the plant kingdom, and show that Au is the most widely distributed SINE family currently identified in plants. Our results also suggest that Au may have originated before the divergence of monocots and eudicots. Au appears to have been lost from some lineages over the course of evolution, and successfully propagated in other lineages. Two types of Au, designated the *Medicago*-type and *Glycine*-type were found in Fabaceae, and the evolution of Au in Fabaceae regarding these two types was examined in detail.

Materials and methods

Database search

BLAST searches (Altschul *et al.*, 1990) with Au sequences from various species were performed on the DDBJ website (http://www.ddbj.nig.ac.jp) to identify species in which Au SINEs are present. Hits under an expectation threshold value of 1 were manually inspected by aligning with previously identified Au elements. Au SINEs were recognized to be present in species, which had sequences with target site duplications.

Plant materials and PCR amplification

Thirty-six species consisting of three ferns, five gymnosperms, five basal angiosperms, twelve monocots (eight Gramineae species), and eleven eudicots were used to examine the presence of Au by PCR (Table 1). Genomic DNA from these species were extracted as described by Escaravage *et al.* (1998). These were amplified by PCR using two sets of internal primers, AUFW1/AURV1 and AUFW2/AURV2. These primers were constructed from the consensus sequences of *Triticum*, *Nicotiana*, and *Medicago*. Reactions were performed in a total volume of 20 μl containing 40 ng of total DNA, 2 μM each of forward and reverse primers, and 0.5 U of Blend *Taq* DNA polymerase (Toyobo, Osaka, Japan). Cycling conditions were 94 °C for 2 min 30 s, and 30 cycles of 94 °C for 30 s, 54 °C (primer AUFW1/AURV1) or 50 °C (AUFW2/AURV2) for 30 s. Reactions with lower annealing temperatures (the lowest of 50 °C for AUFW1/AURV1, and 45 °C for AUFW2/AURV2) were also performed and these produced many bands of unexpected sizes or smears. PCR products of various sizes were cloned using pGEM-T vector (Promega, Madison, Wis.) and more than one clones of each PCR product were sequenced using the ABI 373OXL sequencer. Au was scored as 'amplified' when sequences of more than one PCR product of expected size were confirmed by cloning and sequencing. Primers MtrFW, MtrRV and GmaFW1, GmaRV1, GmaRV2 were constructed from the consensus sequences of *Medicago* and *Glycine*, respectively. MtrRV and GmaRV1 were designed in the region where a 16 bp deletion was observed in *Glycine*, and GmaRV2 was designed in the 3′ end of *Glycine*, which showed high divergence from Au sequences of other species. Primer set MtrFW/MtrRV was expected to amplify only the *Medicago*-type of Au, whereas GmaFW/GmaRV1 and GmaFW/GmaRV2 were expected to amplify only the *Glycine*-type of Au. Reactions of PCR were performed as described above with the annealing temperature at 55 °C. Positions of each primer are indicated in Figure 1 and the sequences are shown in Table 2.

Cloning of Au SINEs from Asimina triloba

Full-length sequences of Au elements from *Asimina triloba*, a basal angiosperm species in which Au was detected by PCR, were cloned by inverse PCR. Total genomic DNA (100 ng) was digested with 37.5 U of *Hind*III in a total volume of 50 μl, self-ligated with 200 U of T4 ligase (New England Biolabs) in a total volume of 500 μl, and amplified by PCR using primers ATinFW/ATinRV (Table 2). PCR reaction was performed in a total volume of 30 μl containing 0.5 μM of each

Table 1. Plant materials used and results of PCR amplification.

Group	Family	Species	AUFW1 & AURV1	AUFW2 & AURV2	MtrFW1 & MtrRV1	GmaFW1 & GmaRV1	GmaFW1 & GmaRV2
Ferns	Lycopodiaceae	*Lycopodium clavatum* L.	−	−	−	−	−
	Psilotaceae	*Psilotum nudum* (L.) Griseb.	−	−	−	−	−
	Equisetaceae	*Equisetum hyemale* L.	−	−	−	−	−
Gymnosperms	Cycadaceae	*Cycas revoluta* Thunb.	−	−	−	−	−
	Ginkgoaceae	*Ginkgo biloba* L.	−	−	−	−	−
	Cupressaceae	*Chamaecyparis pisifera* (Sieb. et Zuco.) Endl.	−	−	−	−	−
	Podocarpaceae	*Decussocarpus nagi* (Thunb.) de Laubenf.	−	−	−	−	−
	Gnetaceae	*Gnetum gnemon* L.	−	−	−	−	−
Basal angiosperms	Nymphaeaceae	*Nymphaea alba* L.	−	−	−	−	−
	Annonaceae	*Asimina triloba* (L.) Dun.	+	+	−	−	−
	Magnoliaceae	*Michelia compressa* (Maxim.) Sang.	−	−	−	−	−
	Magnoliaceae	*Liriodendron tulipifera* L.	−	−	−	−	−
	Lauraceae	*Laurus nobilis* L.	−	−	−	−	−
Monocots	Acoraceae	*Acorus gramineus* Soland.	−	−	−	−	−
	Iridaceae	*Iris japonica* Thunb.	−	−	−	−	−
	Alliaceae	*Allium cepa* L.	−	−	−	−	−
	Commelinaceae	*Tradescantia ohiensis* Raf.	+	+	−	−	−
	Gramineae	*Oryza sativa* L.	−	−	−	−	−
	Gramineae	*Zizania latifolia* L.	−	−	−	−	−
	Gramineae	*Sinobambusa tootsik* Makino	−	−	−	−	−
	Gramineae	*Saccharum officinarum* L.	+	+	−	−	−
	Gramineae	*Sorghum bicolor* (L.) Moench.	+	+	−	−	−
	Gramineae	*Tripsacum dactyloides* L.	+	+	−	−	−
	Gramineae	*Zea mays* L.	+	+	−	−	−
	Gramineae	*Triticum aestivum* L.	+	+	−	−	−
Eudicots	Rosaceae	*Prunus persica* (L.) Batsch.	−	−	−	−	−
	Fabaceae	*Medicago truncatula* Gaertn.	+	+	+	−	−
	Fabaceae	*Lotus japonicus* Regel	+	+	+	−	−
	Fabaceae	*Glycine max* (L.) Merrill	−	+	−	+	+
	Malvaceae	*Gossypium hirsutum* L.	−	−	−	−	−
	Caricaceae	*Carica papaya* L.	−	−	−	−	−
	Cruciferae	*Raphanus raphanistrum* L.	−	−	−	−	−
	Cruciferae	*Brassica oleracea* L.	−	−	−	−	−
	Cruciferae	*Arabidopsis thaliana* (L.) Heynh.	−	−	−	−	−
	Solanaceae	*Nicotiana benthamiana* Domin	+	+	−	−	−
	Boraginaceae	*Symphytum officinale* L.	−	−	−	−	−

+: Au SINE amplified.
−: Au SINE not amplified.

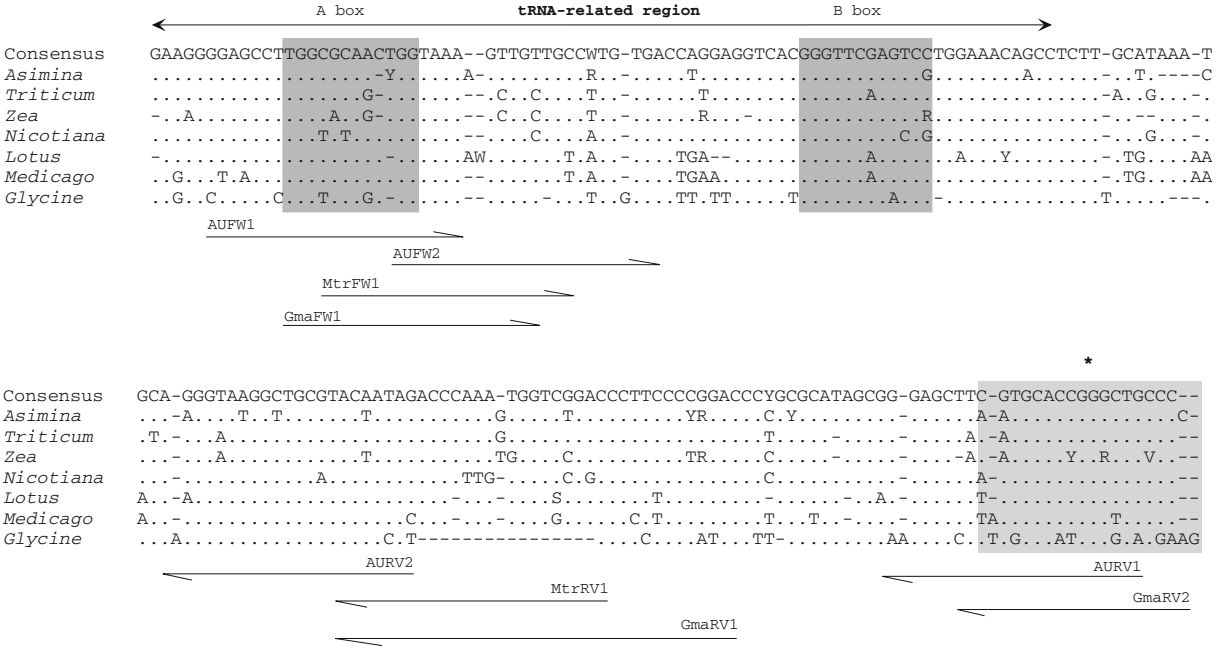

Figure 1. Consensus nucleotide sequences of each species aligned. Dots indicate nucleotides identical to the first sequence, and dashes indicate gaps. 3′ end poly-T tails are not shown, and the shaded region with asterisk was excluded when calculating genetic distances and constructing a phylogenetic tree. A box and B box are also shaded and the tRNA-related region is indicated by arrow. Positions of PCR primers are also indicated by arrows.

primer, and 0.75 U of Blend *Taq* DNA polymerase (Toyobo). Cycling conditions were 94 °C for 2 min 30 s, and 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min, and a final elongation step of 72 °C for 5 min. Resulting products were gel-purified using MAG-EXTRACTOR (Toyobo), then cloned and sequenced as described above. Primers constructed in the flanking regions (ATfl1-FW/RV, ATfl2-FW/RV, ATfl4-FW/RV, Atfl6-FW/RV) were used to amplify full-length sequences by PCR in a total volume of 30 μl containing 20 ng of total genomic DNA of

Table 2. Primers used for PCR amplifications.

		T_m (°C)
AUFW1	5′-GKARCCTTGGCGYARCTGGTAAA-3′	56.0
AURV1	5′-CARCCCGGTGCAYKWAGCTCCC-3′	61.0
AUFW2	5′-TGGTAAAGYTGITGYCWTGTGA-3′	52.1
AURV2	5′-STATWGTACGCAGCCTTWCCCT-3′	55.0
MtrFW1	5′-GCGCAACTGGTAAAGTTGTTGT-3′	55.8
MtrRV1	5′-GTCCCACCATTTGGTGTATTGTA-3′	56.0
GmaFW1	5′-CTGGTGCAGCGGTAAAGTTGT-3′	57.6
GmaRV1	5′-GTATGGGGGAGGGATGTTGTA-3′	57.6
GmaRV2	5′-TTCGTACCCCATTGCCCAGAG-3′	59.5
ATinFW	5′-TGGAAACCACCTCTTGCTTCG-3′	58.5
ATinRV	5′-GACTCGAACCAGTGACCTTCAG-3′	60.4
ATfl1FW	5′-GCTTTATCAAATTGTGGCAGGT-3′	54.8
ATfl1RV	5′-TCCCATATTCTGTATCCCTATGT-3′	57.1
ATfl2FW	5′-ATTCAAATGTAGAGGTGCTGGAA-3′	55.1
ATfl2RV	5′-TAATCAAGGTTTCTCCCACCAAA-3′	55.1
ATfl4FW	5′-AATTGAGCTGTAGCCCCTAGACA-3′	58.7
ATfl4RV	5′-TCCGTAACATCCAACAAATAGCA-3′	55.1
ATfl6FW	5′-CCCTTATGCAATTCCTTCACAAA-3′	55.1
ATfl6RV	5′-AATGTATGGCTTTTCCACTAACAGG-3′	57.2

A. triloba, 0.5 μM of forward and reverse primers, and 1.5 U of *Taq* DNA polymerase (New England Biolabs) at 94 °C for 2 min 30 s, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s. These PCR products were sequenced using the ABI 373OXL sequencer. The resulting nucleotide sequences are deposited in the DDBJ/EMBL/GenBank databases with the accession numbers AB223031-AB223034.

Sequence comparison

Several sequences of Au SINEs were identified in *Triticum*, *Zea*, *Nicotiana*, *Medicago*, *Lotus*, and *Glycine* by database searches. Sequences from these species obtained from the database were aligned together with the sequences of Au SINEs that were cloned from *A. triloba*. Alignment was done using CLUSTAL W v2.0 (Thompson *et al.*, 1994) with some manual refinement (alignment files are available on request). The consensus sequences of each genus were created by assigning the most frequent nucleotide at each position. To reduce the number of sequences, those from the database with indels of five or more nucleotides in comparison to each consensus sequences, and those without recognizable target site duplications were excluded from the analyses. This still left far too many sequences of *Medicago* and *Glycine*, so sequences of these two genera which came under the GSS (genome survey sequence) and HTG (high throughput genomic sequences) division of the DDBJ website were also excluded. The 3′ end region of *Glycine* in particular was difficult to align and therefore the 20 nucleotides at the 3′ end shown in Figure 2 and the poly-T tails were excluded from all sequences in further analyses. An unrooted phylogenetic tree was constructed based on this alignment by the neighbor-joining method (Saitou and Nei, 1987) with Tajima-Nei's distance (Tajima and Nei, 1984) using PAUP v4.0b10 (Swofford, 2002). Bootstrap values were calculated as percentages out of 1000 replicates. Accession numbers of each sequence are given in Appendix 1 of Supplementary Material online.

Results

We first searched the DDBJ database and found Au to be present in the species listed in Table 3. Many Au SINE sequences were found in various species of Gramineae and Solanaceae as expected, and also in species of Fabaceae. No sequences were found in *Oryza sativa* or *Arabidopsis thaliana*.

Next, we examined a broad range of plants for the presence of Au by PCR. Results of PCR are summarized in Table 1. PCR products of unexpected size were observed in many cases, especially when reactions were performed with low annealing temperatures, but sequences of Au could not be found in any of these products. We were able to detect Au in *A. triloba* (Annonaceae) but not in other basal angiosperms, ferns, or gymnosperms. In monocots, Au was detected in *Tradescantia ohiensis* (Commelinaceae), and in some Gramineae species, namely *Zea mays* and *Triticum aestivum*, which had been previously identified, and also in *Tripsacum dactyloides*, *Sorghum bicolor*, and *Saccharum officinarum*. Au could not be identified in other Gramineae species, *Oryza sativa*, *Zizania latifolia*, or *Sinobambusa tootsik*. Eudicot species in which Au were identified by database searches such as *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, and *Nicotiana benthamiana*, all produced PCR products, but Au could not be detected in other eudicots. Both primer sets (AUFW1/AURV1 and AUFW2/AURV2) produced identical results with species apart from *G. max*, in which Au elements were amplified with AUFW2/AURV2, but not with AUFW1/AURV1.

We did come across a large number of sequences annotated as 'retroposon Au-like element' by BLAST searches in many *Brassica* species including *Brassica oleracea*, which have been deposited in GenBank by another group. However, we failed to detect Au SINEs in *B. oleracea* by PCR. The sequences of these 'Au-like elements' are nearly identical to the Au SINEs found in *Triticum* and *Aegilops* species, which would have most likely been detected by our PCR methods had they been present in multiple copies. We also searched the *B. oleracea* genomic database at TIGR (http://www.tigr.org/), which represents approximately one-third of the whole genome, but could not detect a single Au SINE. This is very unlikely for a repetitive element like Au SINE, which appears to be randomly distributed throughout the genome at least in wheat (Yasui *et al.*, 2001), and therefore we conclude that Au is most likely to be absent from *Brassica* species.

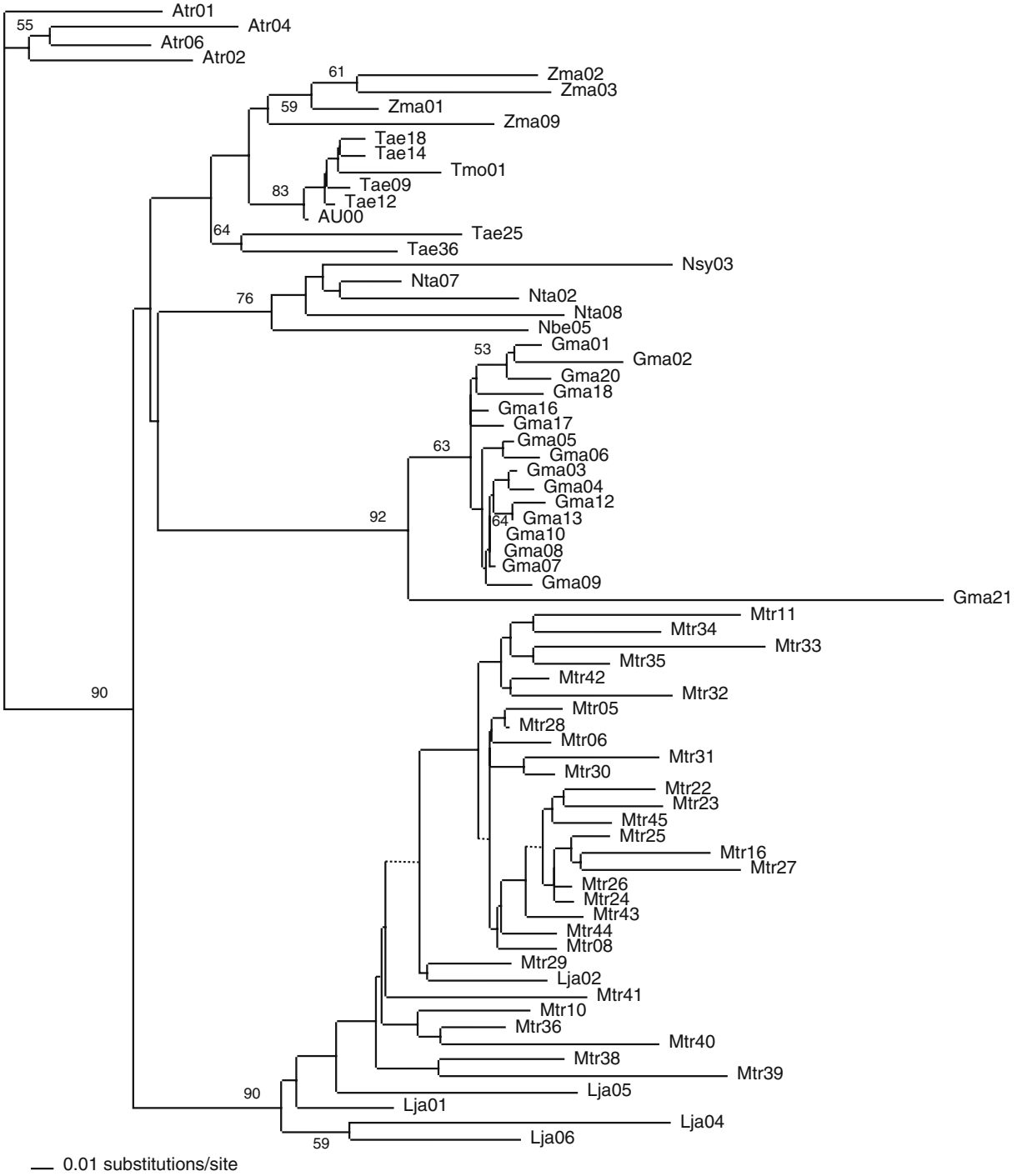

Figure 2. Neighbor-joining tree of sequences of Au SINEs from various species based on Tajima-Nei's genetic distance, with bootstrap values of over 50% indicated. Accession numbers of each sequence are shown in Supplementary Table 1. Abbreviations used are: Atr, *A. triloba*; Tae, *T. aestivum*; Tmo, *T. monococcum*; Zma, *Z. mays*; Nbe, *N. benthamiana*; Nsy, *N. sylvestris*; Nta, *N. tabacum*; Lja, *L. japonicus*; Mtr, *M. truncatula*; Gma, *G. max*. AU00 is the first Au SINE identified from *Aegilops umbellulata*.

Table 3. Species with Au SINEs identified by database searches.

Group	Family	Species
Monocots	Gramineae	*Aegilops speltoides*
		Aegilops tauschii
		Hordeum vulagare
		Leymus chinensis
		Triticum aestivum
		Triticum monococcum
		Triticum turgidum
		Zea mays
Eudicots	Fabaceae	*Glycine max*
		Lotus japonicus
		Medicago sativa
		Medicago truncatula
	Solanaceae	*Capsicum annuum*
		Lycopersicon esculentum
		Nicotiana benthamiana
		Nicotiana plumbaginifolia
		Nicotiana tabacum
		Nicotiana sylvestris
		Solanum bulbocastanum
		Solanum demissum
		Solanum tuberosum

Four full-length sequences of Au from *A. triloba*, each flanked by target site duplications, were successfully cloned. These sequences, together with the sequences of *Triticum*, *Zea*, *Nicotiana*, *Medicago*, *Lotus*, and *Glycine* from the database were aligned and consensus sequences of each genus were created. These showed that all Au sequences of *Glycine* have a 16 bp deletion in their tRNA-unrelated region, and that their 3′ end sequence is distinctive (referred to as '*Glycine*-type' hereafter). Au elements in *Asimina*, *Triticum*, *Zea*, *Nicotiana*, *Medicago*, and *Lotus* do not have a 16 bp deletion, and the 3′ end is nearly identical among these species (referred to as '*Medicago*-type' hereafter). It must be noted that although many sequences with indels or without target site duplications were excluded, no *Glycine*-type sequences of Au were found in *Medicago* or in any other species, and no *Medicago*-type sequences of Au were found in *Glycine* by database searches.

The presence of the *Glycine*-type and *Medicago*-type of Au were also examined by PCR using specific primer sets, which should only amplify either the *Medicago*-type or the *Glycine*-type of Au. MtrFW/MtrRV produced PCR products in *M. truncatula* and *L. japonicus*, but not in *G. max*, whereas GmaFW1/GmaRV1 and GmaFW1/GmaRV2 produced PCR products in *G. max*, but not in *M. truncatula*, *L. japonicus*, or in any other species (Table 1). AURV1 was designed in the 3′ end from the consensus sequences of *Triticum*, *Nicotiana*, and *Medicago*, and AUFW1/AURV1 did not amplify Au in *G. max* either.

An unrooted phylogenetic tree was constructed based on the alignment of Au sequences from *Asimina*, *Triticum*, *Zea*, *Nicotiana*, *Medicago*, *Lotus*, and *Glycine*. These sequences formed well-supported clades of *Asimina*, *Triticum*/*Zea*, *Nicotiana*, *Glycine* and *Medicago*/*Lotus* (Figure 2), suggesting that Au has propagated in each lineage after their divergence. Sequences of closely related species such as *Triticum* and *Zea*, or *Medicago* and *Lotus*, appeared to be closely related as expected, but the Au sequences of *Glycine* did not show any close relationship to the sequences of *Medicago* and *Lotus*.

Discussion

The Au SINE was identified in various angiosperm species by database search and PCR amplification. This classifies Au as the most widespread SINE family currently identified in plants. However, this does not necessarily represent a complete distribution due to limitations in our procedures to detect Au SINEs. For example, although two sets of internal primers were used, and these primers were sufficient to detect Au in species that were identified by database searches, there still remains a possibility that we may have failed to detect Au in certain species due to sequence divergence or indels in the primer regions. The genomic sequences of plant species available in GenBank are very limited, and it is likely that Au SINE will be found in other species as sequences of more species become available. Therefore, we cannot fully conclude that Au is absent from any of the species analyzed apart from fully sequenced species such as *O. sativa* and *A. thaliana*. On the other hand, we were able to confirm the presence of full-length sequences of Au in many species by database searches and also in *A. triloba* by cloning full-length sequences, and we have focused on the evolution of Au using the sequences of these species.

We examined the evolution of Au in Gramineae and Fabaceae in more detail. It has been estimated that Gramineae radiated around 50 MYA (million years ago), forming clades such

as Pooideae, Panicoideae, Oryzoideae, and Bambusoideae (Kellogg, 1998; Paterson *et al*., 2004). It was previously shown that Au is present in many species of Pooideae (*T. aestivum*, *Hordeum vulgare*, *Avena sativa*, and *Sereale secale*) (Yasui *et al*., 2001). In this present study, Au was detected by PCR from species of Panicoideae (*Z. mays*, *T. dactyloides*, *S. bicolor*, and *S. officinarum*) but not in *O. sativa*, *Z. latifolia* (both Oryzoideae), or *S. tootsik* (Bambusoideae). Sequence analyses showed that the sequences of *Triticum* and *Zea* are closely related. This suggests that the Au SINE was lost in the lineage leading to rice after its divergence from the common ancestor of maize and wheat, or in other words within the last 50 million years. At the same time, Au has survived and amplified in many species of Pooideae and Panicoideae, while retaining high sequence similarity as shown with *Triticum* and *Zea*.

In Fabaceae, Au SINE was found in *Medicago*, *Lotus*, and *Glycine*. It is thought that *M. truncatula*, *L. japonicus*, and *G. max* diverged around 40–50 MYA, and that *M. truncatula* shares a more recent ancestry with *L. japonicus* than with *G. max* (Doyle and Luckow, 2003; Choi *et al*., 2004). Choi *et al*. (2004) pointed out that the conservation between the genomes of *M. truncatula* and *G. max* is significantly lower than that of *M. truncatula* and *L. japonicus*. Since the *Medicago*-type Au is present in both *M. truncatula* and *L. japonicus* showing high conservation, and that it shares similar features with Au elements of other lineages, it is likely that it was present in the common ancestor of *M. truncatula*, *L. japonicus*, and *G. max*. However, the *Medicago*-type Au was not found in *G. max*. One possibility is that the *Medicago*-type was simply lost from the genome of *G. max* since its split from *M. truncatula* and *L. japonicus*, as observed in other species such as rice. Considering that Au appears to have been lost from the genome of rice within the last 50 million years, this speculation does not seem unreasonable. Lenoir *et al*. (2005) recently reported that no SINEs were found to be present in orthologous loci of *A. thaliana* and *B. oleracea*, suggesting a high turnover rate of SINEs in these two species. Gene conversion may also have been involved by replacing *Medicago*-type Au elements with *Glycine*-type Au elements. Gene conversion events have been documented in several retroelements and are supposed to occur frequently and have contributed to the sequence evolution of Alu SINEs in primates including humans (Roy *et al*., 2000; Batzer and Deininger, 2002). We did compare the flanking sequences of Au elements of *G. max* with those of *M. truncatula* and *L. japonicus* obtained from the database, but were not able to find any Au insertions in orthologous loci, and therefore cannot confirm the occurrence of gene conversion at this point. The genome of *M. truncatula* is currently being sequenced and comparative maps among Fabaceae species are being constructed. These efforts should enable us to understand more of the evolution of Au SINE in Fabaceae, especially in *G. max*.

The *Glycine*-type Au which has a deletion in its tRNA-unrelated region and a distinctive 3′ end was found only in *G. max* and not in other species including *M. truncatula* and *L. japonicus*. Different 3′ end tails in different species have also been observed in CORE-SINEs and V-SINEs. This has been associated with different 'partner-LINEs' being responsible for their amplification (Gilbert and Labuda, 2000; Ogiwara *et al*., 2002). The amplification of SINEs is dependent on the replicative machinery of LINEs that share a similar 3′ end sequence, and inactivation of these so called partner-LINEs could result in the inactivation of the corresponding SINEs (Kajikawa and Okada, 2002; Dewannieux *et al*., 2003). In such cases, SINEs would need to recruit the 3′ end of another LINE in order to retain its activity. This may be the case with Au SINE in *G. max*. Identifying the LINE family responsible for the amplification of Au should enable us to test this possibility.

Our main aim was to explain the discontinuous distribution of Au, and we were able to confirm the presence of Au in *A. triloba* (Annonaceae). Annonaceae belongs to Magnoliales, which is phylogenetically placed at a basal position to monocots and eudicots (Soltis *et al*., 1999; Sauqet *et al*., 2003). This raises the possibility that Au may have been present before the divergence of monocots and eudicots, which is estimated to be about 150 MYA (Chaw *et al*., 2004). The distribution of Au SINE observed is still confined to certain lineages at this point, and horizontal transfer should be considered as an alternative explanation. However, Au SINE was identified in multiple lineages – *A. triloba*, *T. ohiensis*, Gramineae, Fabaceae, and Solanaceae, and although horizontal transfer would have had to occur several times to explain this distribution, we could not find any evidence indicating the occurrence

of horizontal transfer. The phylogenetic tree constructed using Au sequences placed sequences of closely related species (*Triticum* and *Zea*, *Lotus* and *Medicago*) together, and sequences of distantly related species formed separate clades. We did not observe any case where sequences of distantly related species clustered together, which would be expected if horizontal transfer had occurred. Previous studies also suggest that SINEs and LINEs are generally transmitted vertically, and that horizontal transfers of these retroelements are rare events (Malik *et al.*, 1999; Kajikawa and Okada, 2002). It is likely that Au was lost from the ancestral genome of rice, and it could have also been lost from the genomes of other species, which could explain the limited distribution. Although it is difficult to strictly rule out the possibility of horizontal transfer, we suggest that it is more likely that Au was present in the common ancestor of monocots and eudicots, and lost from many lineages, as observed with *O. sativa*.

This is the first report of a SINE family present in multiple lineages in plants. Au SINE is found in *Z. mays* and *M. truncatula*, whose genome sequencing project is underway, and also in other well-studied species such as *T. aestivum*, *N. benthamiana*, *L. japonicus*, and *G. max*. As more data of these species become available, we should be able to understand more of how this SINE family evolved in each species, and how it contributed to the evolution of these species.

Acknowledgements

We thank Professor O. Ohnishi for revising this manuscript, and two anonymous reviewers for their helpful comments. We also thank Drs. T. Furumoto, S. Hata, T. Nakazaki, Y. Okumoto, Y. Takano and K. Toyoda for providing plant and DNA materials. This work was supported in part by a grant from the Ministry of Education, Culture, Sports, Science and Technology, Japan (Exploratory Research No. 16657027).

References

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.

Batzer, M.A. and Deininger, P.L. 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3: 370–379.

Chaw, S.M., Chang, C.C., Chen, H.L. and Li, W.H. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 58: 424–441.

Choi, H.K., Mun, J.H. and Kim, D.J. et al. (12 co-authors) 2004. Estimating genome conservation between crop and model legume species. Proc. Natl. Acad. Sci. USA 101: 15289–15294.

Dewannieux, M., Esnault, C. and Heidmann, T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35: 41–48.

Doyle, J.J. and Luckow, M.A. 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol. 131: 900–910.

Escaravage, N., Questiau, S., Pornon, A., Doche, B. and Taberlet, P. 1998. Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. Mol. Ecol. 7: 975–982.

Gilbert, N. and Labuda, D. 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc. Natl. Acad. Sci. USA 96: 2869–2874.

Gilbert, N. and Labuda, D. 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. 298: 365–377.

Kajikawa, M. and Okada, N. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111: 433–444.

Kazazian, H.H. Jr. 2004. Mobile elements: drivers of genome evolution. Science 303: 1626–1632.

Kellogg, E.A. 1998. Relationships of cereal crops and other grasses. Proc. Natl. Acad. Sci. USA 95: 2005–2010.

Lenoir, A., Pelissier, T., Bousquet-Antonelli, C. and Deragon, J. M. 2005. Comparative evolutionary history of SINEs in *Arabidopsis thaliana* and *Brassica oleracea*: evidence for a high rate of SINE loss. Cytogenet. Genome Res. 110: 441–447.

Malik, H.S., Burke, D.W. and Eickbush, T.H. 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16: 793–805.

Ogiwara, I., Miya, M., Ohshima, K. and Okada, N. 2002. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res. 12: 316–324.

Paterson, A.H., Bowers, J.E. and Chapman, B.A. 2004. Ancient polyploidization predating divergence of the cereal, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. USA 101: 9903–9908.

Roy, A.M., Caroll, M.L., Nguyen, S.V., Salem, A.H., Oldridge, M., Wilkie, A.O., Batzer, M.A. and Deininger, P.L. 2000. Potential gene conversion and source genes for recently integrated Alu elements. Genome Res. 10: 1485–1495.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Sauquet, H., Doyle, J.A., Scharaschkin, T., Borsch, T., Hilu, K.W., Chatrou, L.W. and Le Thomas, A. 2003. Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. Bot. J. Linn. Soc. 142: 125–186.

Schmid, C.W. 1998. Does SINE evolution preclude Alu function? Nucleic Acids Res. 26: 4541–4550.

Soltis, P.S., Soltis, D.E. and Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402: 358–359.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tajima, F. and Nei, M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269–285.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.

Weiner, A.M. 2002. SINEs and LINEs: the art of biting the hand that feeds you. Curr. Opin. Cell Biol. 14: 343–350.

Yasui, Y., Nasuda, S., Matsuoka, Y. and Kawahara, T. 2001. The Au family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. Theor. Appl. Genet. 102: 463–470.