Characterization of *Soymar1*, a *Mariner* Element in Soybean

Tyler Jarvik and Karl G. Lark

Department of Biology, University of Utah, Salt Lake City, Utah 84112

Manuscript received January 12, 1998
Accepted for publication March 19, 1998

ABSTRACT

Mariner elements, a family of DNA-mediated transposable elements with short, inverted terminal repeats, have been reported in a wide variety of arthropods, as well as planarians, nematodes, and humans. No such element has been reported in a plant. Here we report a *mariner* element in the plant soybean (*Glycine max* (L.) Merr.). Although this sequence belongs to the *mariner* family, it is clearly distinct from previously reported *mariner*-like elements, as well as from the *Tc1* transposon family. Novel aspects of its sequence could be useful as a starting point to identify *mariner*-like elements in new organisms, and it may prove useful in creating a transformation vector for plants.

TRANSPOSABLE elements, pieces of selfish DNA present in the genomes of most organisms, have been grouped into two main categories: retrotransposons, which transpose by means of an RNA intermediate; and DNA mediated transposons, such as *mariner, Tc1*, and the Drosophila *P* element (Berg and Howe 1989). In plants, in which genetic manipulation would be most useful, DNA-mediated transposable elements of the inverted terminal repeat (ITR) type are of particular interest in transformation and transposon tagging of genes (*e.g.*, Osborne and Baker 1997; Bhatt *et al.* 1996; Cooley *et al.* 1996). *Mariner* elements do not rely on species-specific host factors (Lampe *et al.* 1996), and they have shown promise as transformation vectors even in organisms that were not the source of the element (Gueiros-Filho and Beverly 1997). Although *mariner* elements are known to spread by horizontal transfer (Capy *et al.* 1994; Garcia-Fernàndez 1995), *mariner*- and *Tc1*-like elements have not yet been discovered in plants.

Mariner-like elements are widespread in animals occurring in several phyla at frequencies anywhere between two copies and 10,000 copies per haploid genome (Capy *et al.* 1992; Garcia-Fernàndez *et al.* 1995). They are relatively small, typically ~1300 bp, and flanked by a short (30–40-bp) ITR sequence (Jacobson *et al.* 1986). They contain a single open reading frame encoding a transposase of ~330 amino acids, which is capable of recognizing the ITR sequence, excising the transposon, and inserting it at a target sequence, the dinucleotide TA, elsewhere in the genome (Lampe *et al.* 1996). As a result of this mobility, copies of the transposon are dispersed throughout the genome, resulting in numerous polymorphisms that will tend to segregate independently in a genetic cross. If a probe corresponding to part of such an element is used to detect restriction fragments, dominant polymorphisms may be detected wherever the transposon has been inserted. Genetic elements that are not mobile are unlikely to produce such a pattern.

As part of a project to create a molecular map of the soybean genome, we isolated several hybridization probes that suggested the existence of such transposable elements (Lark *et al.* 1993a; Cregan *et al.* 1998). The sequence of one of these is very similar to sequences of *mariner* elements from other organisms. Its characteristics are described below.

Corresponding author: Karl G. Lark, University of Utah, Department of Biology, 257 South 1400 East, Salt Lake City, UT 84112-0840. E-mail: lark@bioscience.utah.edu

MATERIALS AND METHODS

Plant lines and RFLP analysis: The G214 clone was isolated from a λ library (prepared by Dr. R. Goldberg) as described by Frazier (1988). Soybean DNA from the cultivar 'Minsoy' (PI 27890) was obtained as described previously (Lark *et al.* 1993a) from plants grown in the greenhouse using seed obtained from R. Nelson (University of Illinois, Chicago). Southern transfers were prepared from this DNA and hybridized to a radioactive DNA probe prepared from the G214, as described by Lark *et al.* (1993a). The G214 probe also was hybridized to a southern transfer made from a *Taq*I digest of genomic DNA from 24 unrelated soybean lines (Lark *et al.* 1992, 1993b; Figure 1).

Sequencing: Subclones of G214 were sequenced manually using the dideoxy chain termination method. After initial sequencing revealed the presence of an ITR, a primer was designed from the ITR sequence, and additional primers were designed from the sequence internal to the ITRs. The ITR primer was used to PCR amplify other sequences from the soybean genome using standard methods. To amplify longer PCR fragments, we used the method of Barnes (1994). Long-range PCR reactions (50 μl) contained 1× buffer PC2 (AB Peptides, St. Louis, MO), 1.5 μm primer, 250 μm dNTP, 20 ng genomic DNA from 'Minsoy,' and 0.8 μl of a 16:1 mixture of Klentaq1 polymerase (AB Peptides) and *Pfu* polymerase. Reactions were amplified for 40 cycles of 94° for 20 sec and 68° for 10 min. PCR products were separated from each other

on 1% agarose gels, purified with a QiaQuick gel extraction kit (Qiagen, Chatsworth, CA), and sequenced at the University of Utah Core Sequencing Facility with a sequencer (model 377; Applied Biosystems Inc., Foster City, CA), using the internal primers and other primers designed from the previous sequencing run.

Sequence analysis: Nucleotide and amino acid sequence searches were carried out using the GCG Wisconsin Package version 9.0 (Genetics Computer Group, Madison, WI) and the Baylor College of Medicine (BCM) Search Launcher web site (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). Searches were made using TFASTA, BLASTP, and WU-BLASTP, using default parameters. Dot plots were produced using the GCG commands COMPARE and DOTPLOT, using a window size of 30 and a stringency of 20. Alignments were made with the GCG command PILEUP, using the PAM250 scoring matrix, a gap weight of 2.0, and a gap length weight of 0.05. Alignments were edited slightly by eye using the Macintosh program Seq-Vu.

RESULTS

Genomic polymorphism: Using several restriction enzymes and three populations of recombinant inbred segregants derived from genetic crosses between three parents, the G214 probe identified restriction fragments that mapped to at least 25 distinct positions on 13 of the 20 soybean chromosomes (Cregan *et al.* 1998). Several fragments, which did not exhibit polymorphism in any of the three segregant populations, could represent additional positions on other chromosomes. Because fragments with similar map positions were identified on blots made with different restriction enzymes, it was difficult to determine if each of these positions represents a single copy or a cluster of two or more copies. However, in at least two cases, two distinct fragments identified by the same enzyme mapped close together (within 3 cM). The G214 probe was also hybridized to a southern transfer of 24 unrelated soybean lines, mostly plant introductions (Figure 1). As can be seen, the element identified by G214 was found to be present in each of the lines in approximately equal numbers, but in very different positions.

Sequencing: The clone for the G214 probe was found to contain two 40-bp inverted repeats that were 95%

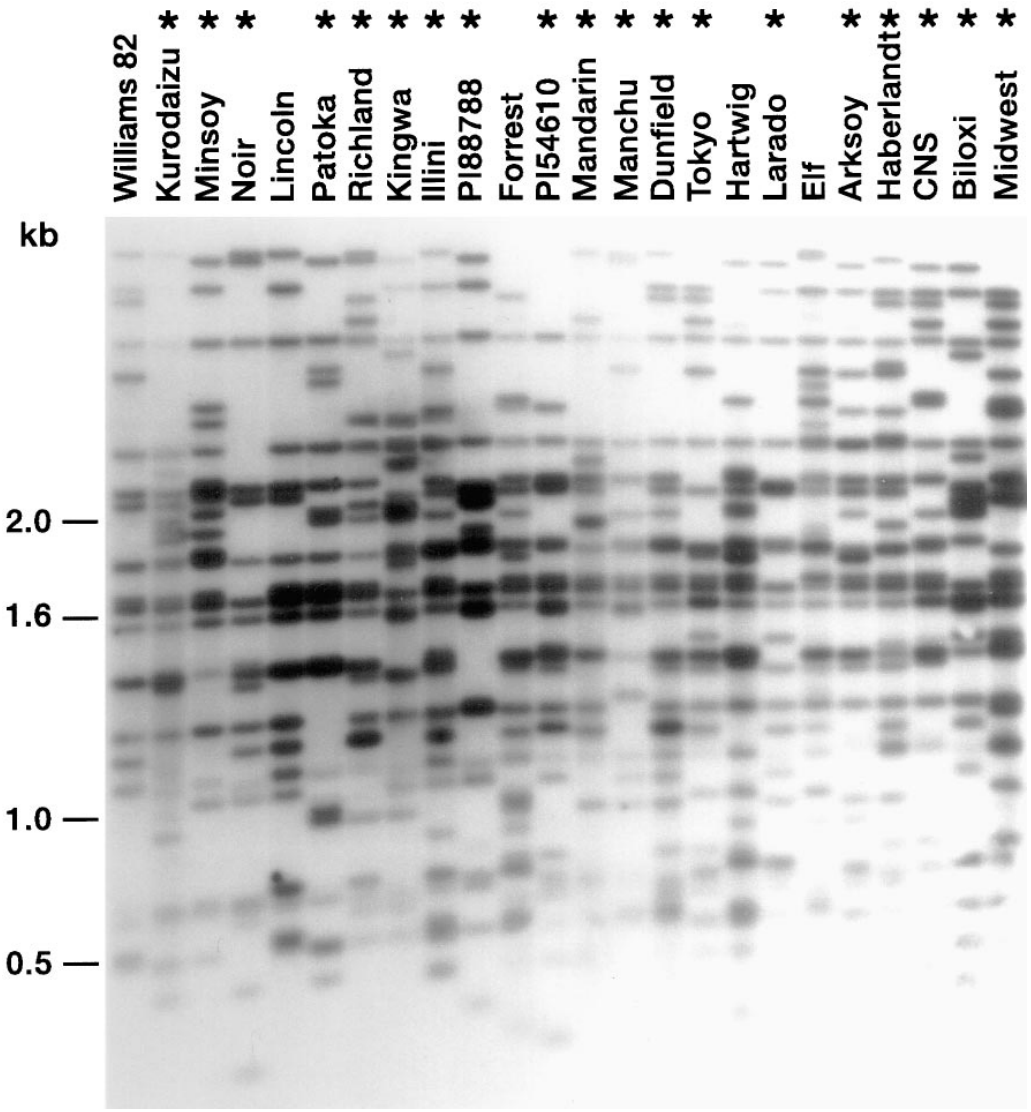

Figure 1.—Positions of G214 in various soybeans. Southern blot of a *Taq*I digest of genomic DNA from 24 unrelated soybean lines that was probed with the original 0.5-kb G214 element. Plant introductions (no record of crossbreeding) are noted with an asterisk.

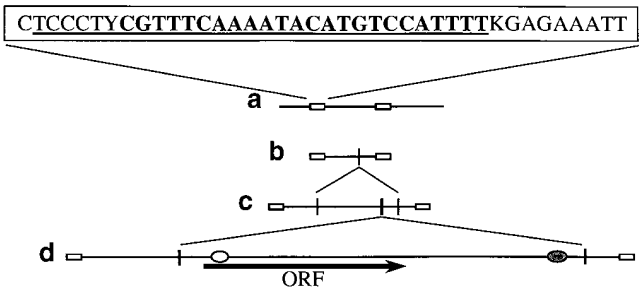

Figure 2.—Positions of ITR sequences, amplification primers, and ORFs in the original G214 clone (a) and PCR products (b–d). Rectangles indicate ITR sequences. The open and filled ovals represent the primers used for amplification of the 2.0-kb sequencing template (see results). The complete ITR sequence is shown above (a). Ambiguity codes (Y for C or T, K for G or T) are used where the left and right repeat did not agree. Bold type represents the primer used to amplify the fragments in b and c. The underlined sequence, degenerate at one position, is the primer used to amplify the fragment in d.

identical and separated by 425 bp (Figure 2a). To the outside of each ITR was the dinucleotide TA. We hypothesized that the G214 probe represented a "gutted" copy of the complete transposon capable of hybridizing to, but not necessarily identical to, other copies of the transposon in the genome. Therefore, in the hopes of amplifying the complete element, we selected a single primer from the ITR. PCR amplification of soybean genomic DNA using this primer produced a product of the expected 500-bp size (Figure 2b), as well as one of ~1 kb (Figure 2c). This last one was found to be nearly identical to the smaller copy, but contained a 456-bp insertion beginning at bp 318. Using a somewhat longer ITR primer and long-range PCR conditions, we then produced an additional 3.5-kb product (Figure 2d). This PCR product was sequenced directly using primers derived from the sequence of the 1-kb copy. The 3.5-kb sequence so derived was found at each end to be nearly identical to the 1-kb copy, with a 2.5-kb insertion beginning at bp 790. Enough of the 2.5-kb insertion was sequenced to design a set of primers specific to the unique region (ovals in Figure 2d). Using these primers, PCR amplification from genomic DNA and from the 3.5-kb fragment both produced a single product of the expected 2.0-kb size. The 2.0-kb fragment amplified from the genomic DNA was sequenced directly. The sequence at the ends of this fragment was identical to that derived from the 3.5-kb product, confirming that these were contiguous. When the complete sequence was compiled, it was found to contain a 425-aa open reading frame beginning at bp 993 (Figure 2d).

Sequence analysis: Using the WU-BLASTP program, a match was found between the amino acid translation of the ORF and a planarian *mariner* transposase with a sum $P(2)$ score of 2.6×10^{-7}. A dot plot of these two sequences is shown in Figure 3. Two regions of similarity were identified by WU-BLASTP, an algorithm that allows gapped alignments. The WU-BLASTP search also generated good matches with several other transposases, including the human *mariner* transposase (P score 1.8×10^{-5}) and the Drosophila *S* element (P score 2.0×10^{-4}). Other search algorithms, such as TFASTA, found similar matches,

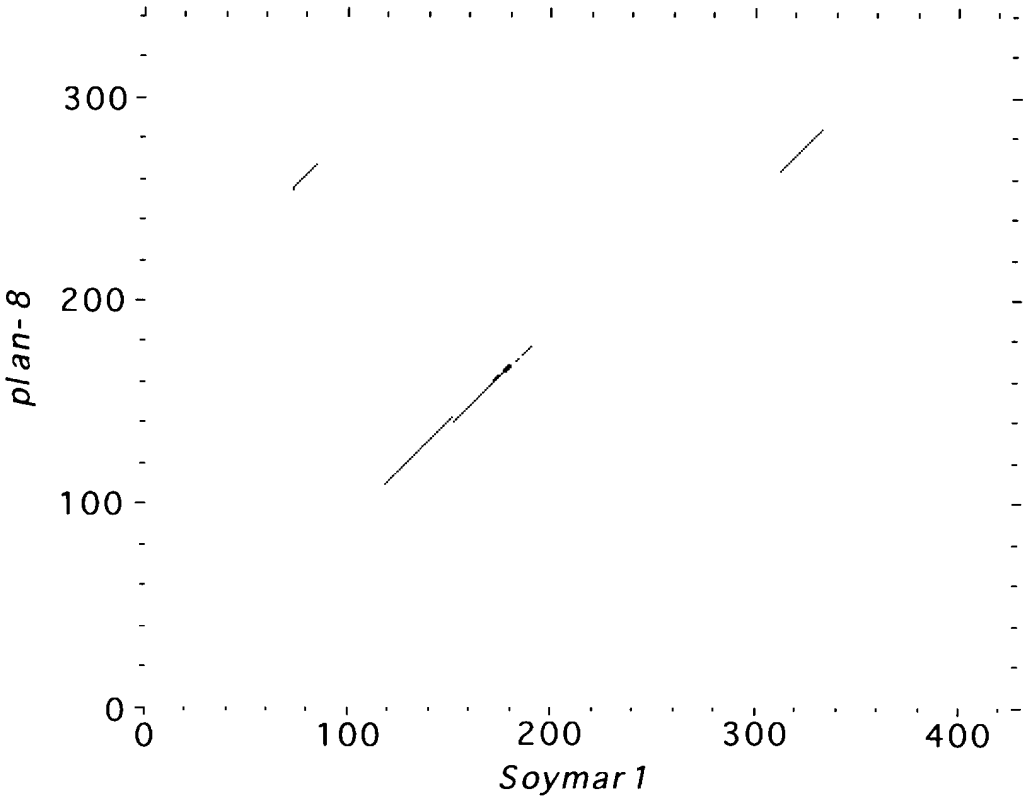

Figure 3.—Dot plot of the *Soymar1* amino acid sequence compared to the amino acid sequence of the planarian *mariner-8* transposase (plan-8). Comparison used a stringency of 20 and a window size of 30 (see materials and methods). Two regions of similarity are seen on the main diagonal, plus one region of probable noise above the diagonal. The right-hand region on the main diagonal corresponds to the D35E region.

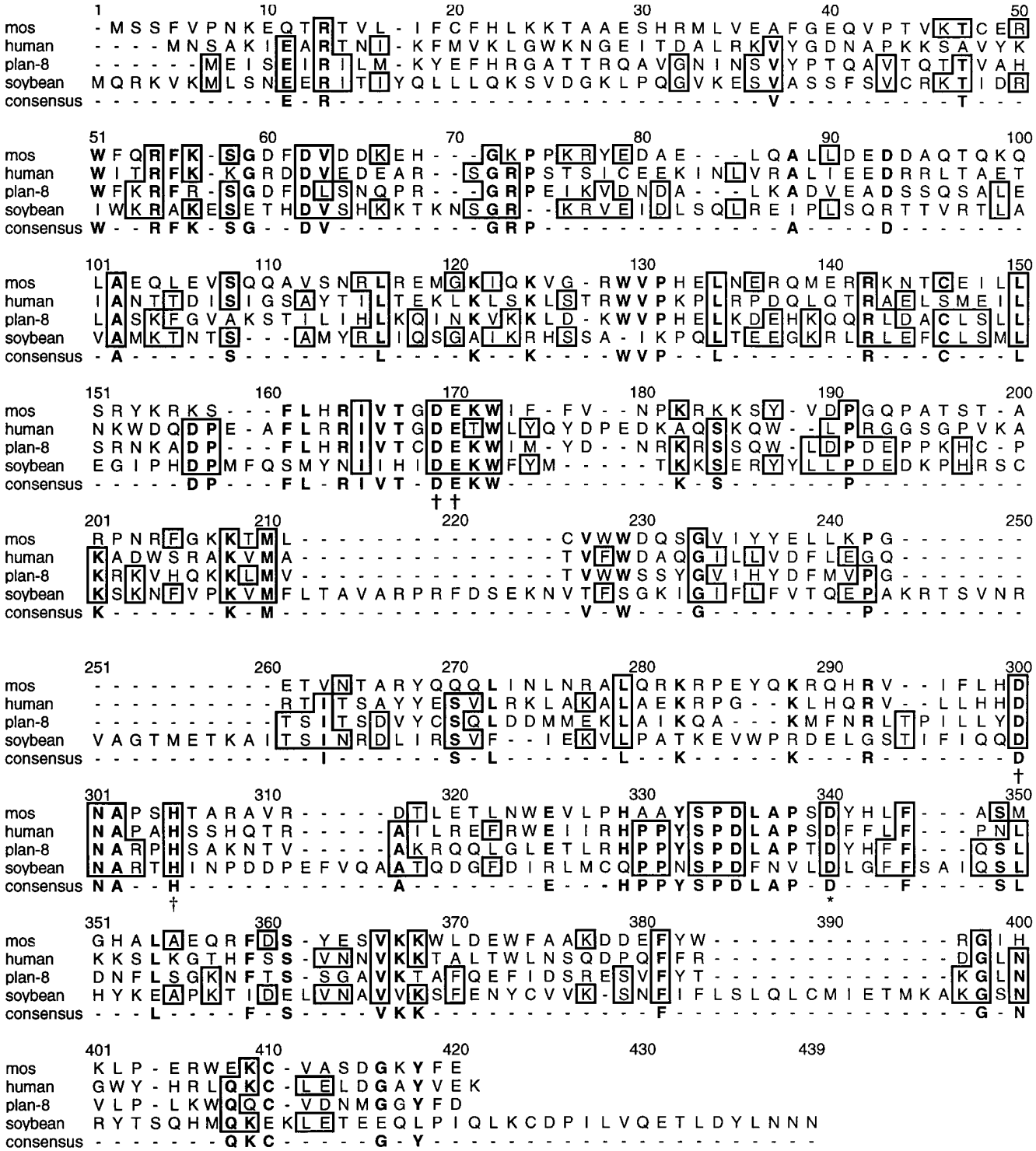

Figure 4.—Alignment of the *Soymar1* amino acid sequence (soybean) with three other *mariner* transposases: the planarian *mariner-8* transposase (plan-8; GenBank accession no. Q24700), the human *mariner1* transposase (human; GenBank accession no. Q13539), and the translation of the Drosophila *Mos1* gene (mos; GenBank accession no. X78906). The consensus represents identity with a plurality of three. Agreement with the consensus is shown in bold. Residues common to *Soymar1* and one or more other sequences are boxed. The mariner D35D region is located between aa 300 and 340 in the alignment. * The aspartic acid residue (alignment position 340) that is unique to all mariner transposases. † Amino acids (alignment positions 169, 170, 300, and 305) that are completely or prominently conserved in the *IS630-Tc1* family, including *mariners* (Doak *et al.* 1994).

but these matches had lower scores (2.3×10^{-5} for the planarian *mariner*) because only the first region was found.

We compiled an alignment (Figure 4) of the soybean amino acid sequence with three representative *mariner* transposase sequences: the planarian *mariner-8* transposase (which produced the highest scoring match with the soybean sequence), the Drosophila *mos1* transposase (an active transposase; see Medhora *et al.* 1991), and the human *mariner* transposase. From this alignment, it is clear that more than half of the amino acids conserved between these *mariners* are also present in the soybean sequence, including the known catalytic residues (Doak *et al.* 1994; Vos and Plasterk 1994).

DISCUSSION

Our data present the first evidence that a *mariner* element, which we have named *Soymar1*, occurs in a plant. *Soymar1* may be slightly less similar to other *mariners* than most other *mariners* are to each other, but it has identities with other *mariners* at several important positions. The most significant identity is in the well-known "D35E" motif, a motif present in an extremely wide variety of transposons (Doak *et al.* 1994). In *mariner* transposons, the glutamic acid residue of the "D35E" is replaced by a second aspartic acid residue (Robertson 1995). This feature is unique to *mariner* elements and is shared by *Soymar1*. Therefore, unless *Soymar1* is to be considered the first member of its own transposon family, it should be considered a *mariner*. It is certainly divergent enough from previously reported *mariners* to suggest that it represents a new *mariner* subfamily. Because the *Soymar1* sequence was obtained from multiple PCR products rather than from a single clone, it is possible, though unlikely, that some features of the sequence are a result of PCR artifacts. The extent of such artifacts will become evident as examples of the element from other soybeans are sequenced.

Aside from the amino acid sequence differences (including the 30 or so "extra" amino acids in *Soymar1* between aa 200 and 250, and the C-terminal 20 or so amino acids), *Soymar1* also differs from other *mariners* in the sequence of its ITR. A FASTA search using the *Soymar1* ITR sequence turned up no matches to other ITRs, and direct comparison with *mariner* ITRs reveals little similarity. Although the only ITR sequences available are the two from the original clone, it is likely that they are quite similar to the ITRs of other copies of *Soymar1* because they differ from each other by only two nucleotides out of 40, and because the ITR primer designed from them was capable of annealing to the ITRs of the longer elements.

The identification of an element representing either a new *mariner* subfamily or new transposon family will enable researchers to search for related elements in other organisms. Searching for *mariner* elements in new organisms requires the use of a probe or PCR primers designed from regions of conserved sequence (Robertson and MacLeod 1993). *Soymar1* does not share some of the amino acids that are conserved between other mariners; therefore, *Soymar1* and any close relatives it may have would not have been detected by some of the previous assays, including that used by Robertson and MacLeod (1993). Primers or probes designed from the *Soymar1* sequence could be used to identify related elements if they occur in other plants or in other organisms, such as insects, from which they could have been horizontally transmitted to soybean.

Despite the differences between the *Soymar1* transposase and previously described transposases, the putative element is relatively intact. In eukaryotes, most transposons are expected to be inactive because of lack of selection pressure (Kaplan *et al.* 1985), a prediction that is confirmed for most *mariner* elements found to date (reviewed in Lohe *et al.* 1995). The presence of two kinds of "gutted" copies of *Soymar1* (copies containing the ITR but lacking the transposase) is not remarkable, nor is the presumably non-*mariner* insertion found after the ORF in the long copy that we sequenced. However, the fact that the *Soymar1* sequence contains neither stop codons nor frameshifts is quite unusual, and while a frameshift could be counteracted by an artifact of PCR, it is highly unlikely that a PCR artifact would remove existing stop codons, or that PCR artifacts would have removed a large number of defects. Although the presence of defective elements means that the transposon probably was not introduced into the soybean genome very recently, the presence of such a defect-free transposase gene suggests that it is possible that active copies still exist in soybean. PCR assays have revealed the presence of the *Soymar1* ITR in several lines of *Glycine soja*, the annual wild progenitor of soybean, but this sequence has not been detected by similar assays in several species of perennial soybean (data not shown). Thus, the element was probably introduced into soybeans between the time that the annual and perennial species diverged, but before domestication.

The question of whether an active transposon still exists remains unresolved. The examples in Figure 1 are consistent with recent movement of the element because most of the lines shown (those marked with asterisks) are not the result of crossbreeding and soybean is a natural inbreeder. If the *Soymar1* element has not been active since the domestication of soybean ($\sim$3000 years ago), the patterns in Figure 1 could only be the result of crossbreeding or numerous independent domestication events. While it is known that domestic soybean originated in a very small area of northern China, no accurate data are available concerning the number of domestications that may have occurred.

Only one copy of the transposase gene, derived from the soybean cultivar 'Minsoy,' has been sequenced. More than 9000 different soybean or *G. soja* accessions

are available from the U.S. Department of Agriculture germplasm collection (*e.g.*, Bernard *et al.* 1989). PCR amplification of related sequences from other soybeans could produce complete elements without the insertions or deletions found in 'Minsoy.' Even if no active element is found, the transposase gene is probably sufficiently intact that an active transposase could be reconstructed by sequencing copies from a number of different soybeans, including wild *G. soja*. Such an active transposase would be extremely useful in the context of a transformation vector.

We wish to thank Tom Doak and Cynthia Lark for comments and help during the course of this work and in the preparation of the manuscript. We thank Margaret Robertson and the University of Utah Core Sequencing Facility for assistance in the latter stages of DNA sequencing. This work was supported by a grant from the United Soybean Board.

LITERATURE CITED

Barnes, W. M., 1994 PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. Proc. Natl. Acad. Sci. USA **91:** 2216–2220.

Berg, D. E., and M. M. Howe, 1989 Mobile DNA. American Society for Microbiology, Washington, DC.

Bernard, R. L., G. A. Juvik and R. L. Nelson, 1989 USDA Soybean Germplasm Collection Inventory. International Agricultural Publications, Urbana, IL.

Bhatt, A. M., T. Page, E. J. Lawson, C. Lister and C. Dean, 1996 Use of *Ac* as an insertional mutagen in Arabidopsis. Plant J. **9:** 935–945.

Capy, P., J. R. David and D. L. Hartl, 1992 Evolution of the transposable element *mariner* in the Drosophila melanogaster species subgroup. Genetica **86:** 37–46.

Capy, P., T. Langin, Y. Bigot, F. Brunet, M. J. Daboussi *et al.*, 1994 Horizontal transmission versus ancient origin: *mariner* in the witness box. Genetica **93:** 161–170.

Cooley, M. B., A. P. Goldsbrough, D. W. Still and J. I. Yoder, 1996 Site-selected insertional mutagenesis of tomato with maize *Ac* and *Ds* elements. Mol. Gen. Genet. **252:** 184–194.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark *et al.*, 1998 An integrated genetic linkage map of the soybean. Crop Sci. (in press).

Doak, T. G., F. P. Doerder, C. L. Jahn and G. Herrick, 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.

Frazier, B. L., 1988 Restriction fragment length polymorphisms as genetic markers in soybean. M. Sci. Thesis, University of Utah, Salt Lake City.

Garcia-Fernàndez, J., J. R. Bayascas-Ramírez, G. Marfany, A. M. Muñoz-Mármol, A. Casali *et al.*, 1995 High copy number of highly similar *mariner*-like transposons in planarian (platyhelminthe): evidence for a trans-phyla horizontal transfer. Mol. Biol. Evol. **12:** 421–431.

Gueiros-Filho, F. J., and S. M. Beverly, 1997 Trans-kingdom transposition of the Drosophila element *mariner* within the protozoan Leishmania. Science **276:** 1716–1719.

Jacobson, J. W., M. M. Medhora and D. L. Hartl, 1986 Molecular structure of a somatically unstable transposable element in Drosophila. Proc. Natl. Acad. Sci. USA **83:** 8684–8688.

Kaplan, N., T. Darden and C. H. Langley, 1985 Evolution and extinction of transposable elements in Mendelian populations. Genetics **109:** 459–480.

Lampe, D. J., M. E. A. Churchill and H. M. Robertson, 1996 A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. EMBO J. **15:** 5470–5479.

Lark, K. G., J. Evans, F. Basha, R. Bogden, R. Copeland *et al.*, 1992 Molecular phylogeny as a tool for soybean breeding. Soybean Genet. Newslett. **19:** 174–181.

Lark, K. G., J. M. Weissemann, B. F. Mathews, R. Palmer, K. Chase *et al.*, 1993a A genetic linkage of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1.' Theor. Appl. Genet. **86:** 901–906.

Lark, K. G., J. Evans, P. Wilhelm, S. Atkinson, M. Christensen *et al.*, 1993b Molecular phylogeny as a tool for soybean breeding II. Soybean Genet. Newslett. **20:** 197–201.

Lohe, A. R., E. N. Moriyama, D. Lidholm and D. L. Hartl, 1995 Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol. Biol. Evol. **12:** 62–72.

Medhora, M., K. Maruyama and D. L. Hartl, 1991 Molecular and functional analysis of the mariner mutator element *Mos1* in Drosophila. Genetics **128:** 311–318.

Osborne, B. I., and B. Baker, 1995 Movers and shakers: maize transposons as tools for analyzing other plant genomes. Curr. Opin. Cell Biol. **7:** 406–413.

Robertson, H. M., 1995 The *Tc1-mariner* superfamily of transposons in animals. J. Insect Physiol. **41:** 99–105.

Robertson, H. M., and E. G. MacLeod, 1993 Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. **2:** 125–139.

Vos, J. C., and R. H. Plasterk, 1994 *Tc1* transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J. **13:** 6125–6132.

Communicating editor: R. H. Davis