J Mol Evol (1999) 48:684–691

JOURNAL OF MOLECULAR EVOLUTION

Characterization of Repetitive DNA Elements in *Arabidopsis*

Stefan A. Surzycki, William R. Belknap

United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710, USA

Received: 13 October 1998 / Accepted: 30 December 1998

Abstract. We have applied computational methods to the available database and identified several families of repetitive DNA elements in the *Arabidopsis thaliana* genome. While some of the elements have features expected of either miniature inverted-repeat transposable elements (MITEs) or retrotransposons, the most abundant class of repetitive elements, the *AthE1* family, is structurally related to neither. The *AthE1* family members are defined by conserved 5′ and 3′ sequences, but these terminal sequences do not represent either inverted or direct repeats. *AthE1* family members with greater than 98% identity are easily identified on different *Arabidopsis* chromosomes. Similar to nonautonomous DNA-based transposon families, the *AthE1* family contains members in which the conserved terminal domains flank unrelated sequences. The primary utility of characterizing repetitive sequences is in defining, at least in part, the evolutionary architecture of specific *Arabidopsis* loci. The repetitive elements described here make up approximately 1% of the available *Arabidopsis thaliana* genomic sequence.

Key Words: Miniature inverted-repeat transposable element — Retrotransposon — Recombination — Plant

Introduction

Repetitive DNA, in particular mobile genetic elements, represent significant portions of many plant genomes (Bureau et al. 1996; SanMiguel et al. 1996). Mobile elements are classified into two basic groups based upon their mode of transposition (Berg and Howe 1989). DNA transposons, such as the *Ac/Ds* family (Federoff 1989), are mobilized via DNA intermediates, and retrotransposable elements are mobilized through RNA (Weiner et al. 1986). While the evolutionary roles of these elements remain to be clearly defined (Lonnig and Saedler 1997), it is clear that they have the potential to serve as a major source of mutation and to contribute specific regulatory sequences to genes (Britten 1997; Wessler 1996).

Arabidopsis thaliana, an invaluable model species for studies of plant biology, has a relatively low content of repetitive DNA (Bureau et al. 1996; Pruitt and Meyerowitz 1986). A number of transposons (Frank et al. 1997), retrotransposable elements (Chye et al. 1997; Konieczny et al. 1991; Pelissier et al. 1995; Wright et al. 1996), and other repetitive DNAs (Martinez-Zapater et al. 1986; Richards et al. 1991; Schmidt et al. 1995; Simoens et al. 1988; Thompson et al. 1996a–c) have been identified in this organism.

We have previously employed search algorithms based upon identification of inverted repeated domains to characterize repetitive DNA elements from solanaceous plants (Oosumi and Belknap 1997; Oosumi et al. 1995b), *C. elegans* (Oosumi et al. 1995b, Oosumi et al. 1996), and humans (Oosumi et al. 1995a). An alternative algorithm was employed here to identify several repeat families in *Arabidopsis.* One of the subfamilies of repeats is contained within a previously identified repeated domain (repeat ATR0053 of the AtRepBase, N.N. Dedhia and G.P. Copenhaver, Cold Spring Harbor Laboratory). The development of a complete inventory of these elements is important for both an accurate definition of repetitive sequence domains with the database and the

Correspondence to: William Belknap; e-mail: wrb@pw.usda.gov

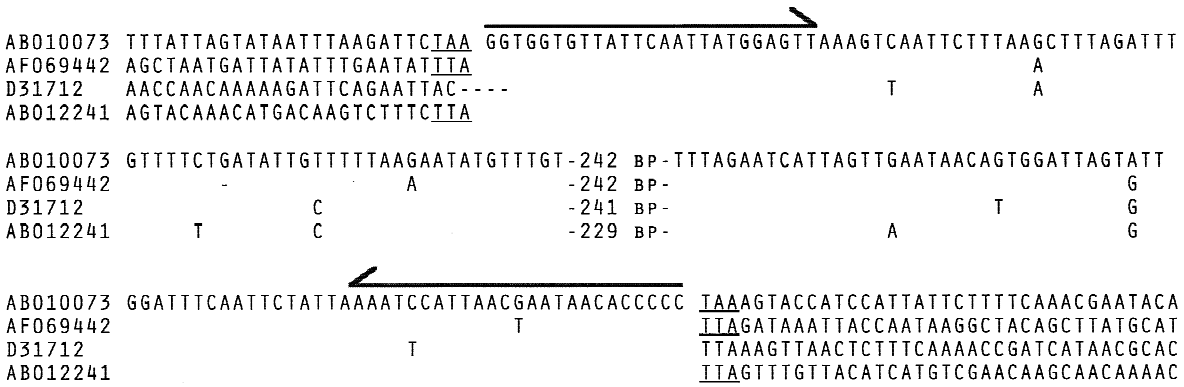

Fig. 1. Alignment of the 5′ and 3′ ends of *MAthE1* repeats from *Arabidopsis.* Inverted-repeat domains are indicated by *arrows* and flanking direct repeats are *underlined.* Sizes of regions of the repeats not shown are indicated. Within the repeated domains, only mismatched bases are shown and gaps are indicated by *dashes.* Positions in GenBank accession numbers for nucleotides shown are as follows: AB010073, 74,004–74,471 (reverse complement); AF069442, 17,128–17,594; D31712, 73–536; and AB012241, 38,431–38,898. The repeated domain in D31712 is located at position −850 relative to the transcription start (Ohta et al. 1995).

characterization of their contribution to the evolutionary architecture of the *Arabidopsis* genome.

Materials and Methods

Computational Analysis

Repeated sequences were identified within the *Arabidopsis thaliana* genome using a search algorithm written in C programming language and run under a Linux platform (Micron Millenia 300MHz PC). *Arabidopsis* sequences for analysis were downloaded via ENTREZ from the NCBI server in Maryland. Both individual BAC clones and approximately 1-Mb segments of a large contig (Bevan et al. 1998) were entered into the search routine. The search algorithm compared similarities between two user-defined windows. Similarity between the sequence windows was either reported or ignored based upon user-defined GC content and match percentage.

When a match with percentage identity *(P)* and GC content above the user threshold was located, a separate algorithm was employed to define the approximate length of the repeated sequence. The original window size *(X)* was increased ($X = X + 1$) simultaneously in both the 5′ and the 3′ directions. With each increase in window size, the percentage identity of the expanded windows (P_e) was determined, expansion was continued until P_e dropped to a value two percentage points less than the value observed in the main algorithm. The output files from individual searches were then screened for related sequences using BLAST Network Service of the National Center for Biotechnology Information (Altschul et al. 1990) using the default matrix (Altschul et al. 1990). The actual boundaries of the repeated domains described here were determined by direct comparison of related repeats from multiple loci using MacVector (Eastman Kodak Co.). AssemblyLIGN (Eastman Kodak Co.) was used in the derivation of consensus sequences.

Results

Repetitive Elements Resembling Miniature Inverted-Repeat Transposable Elements (MITEs)

The computational survey employed here revealed three families of repetitive DNAs resembling MITEs (Bureau et al. 1996; Wessler et al. 1995) in *Arabidopsis.* The first class of repeats, designated *MAthE1,* is small (approximately 400-bp) elements defined by 25-bp imperfect inverted repeats (Fig. 1). The inverted repeats are flanked by a trinucleotide AT-rich direct repeats, representing a potential target site duplication common to these types of elements (Bureau et al. 1996; Bureau and Wessler, 1992). Highly similar copies of *MAthE1* are found on several *Arabidopsis* chromosomes. The locus *ATHCP31C* (Fig. 1; accession D31712) encodes the gene for the chloroplast binding protein cp31 (Ohta et al. 1995). In this case the *MathE1* element is located in the promoter region, 840 bp 5′ to the transcription start site. The *MAthE1* elements are relatively rare in the genome, with fewer than 10 copies in the available database.

The second group of repetitive elements with structural features similar to MITEs is the *MAthE2* family. As shown in Fig. 2, these approximately 600-bp elements are defined by 24-bp imperfect inverted repeats, flanked by potential dinucleotide TA target site duplications (Bureau and Wessler, 1994; Oosumi et al. 1996). The *MathE2* elements are considerably more abundant than the *MathE1* repeats; more than 70 *MAthE2* repeats are easily identified in the available database by using the terminal inverted-repeat domains in Fig. 2 as a query.

The *MAthE2* query also revealed the presence of a *MAthE2*-like element in the tomato polyphenol oxidase A gene (*LepooxA*) (Newman et al. 1993). The portion of the *LepooxA* gene flanked by *MAthE2*-like sequences spans the region from −1694 to −93 relative to the translation start site and includes the putative CAAT box (Newman et al. 1993).

The third family of MITE-like repetitive sequences found are the *MAthE3* elements (Fig. 3). The *MAthE3.1* repeats (Fig. 3A) are defined by 135-bp terminal inverted repeats, are approximately 300 bp in length, and are flanked by 8-bp direct repeats. The *MAthE3.1* elements are less abundant than the *MAthE2* repeats, with 11 cop-

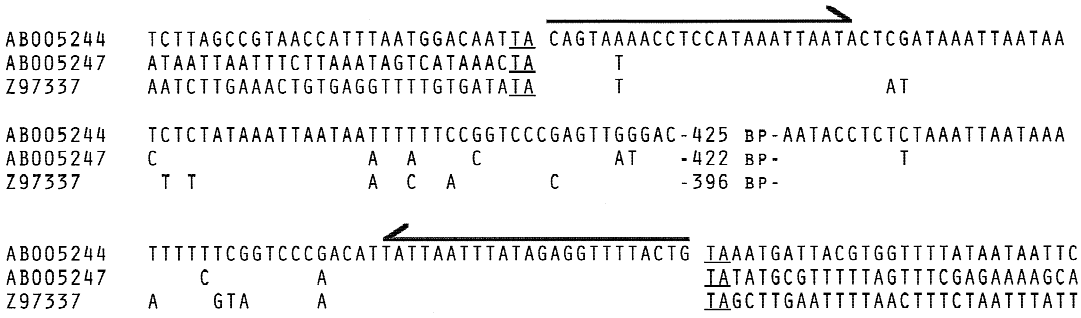

Fig. 2. Alignment of the 5′ and 3′ ends of *MAthE2* repeats from *Arabidopsis.* Inverted and direct repeats, mismatches, and gaps are indicated as in Fig. 1. Sizes of regions of the repeats not shown are indicated. Positions in GenBank accession numbers for nucleotides shown are as follows: AB005244, 417–1045; AB005247, 28,249–28,874; and Z97337, 26,952–27,551.

Fig. 3. Alignment of *MAthE3* repeats from *Arabidopsis.* **A** Alignment of *MAthE3.1* repeats. **B** Alignment of the 5′ and 3′ ends of *MAthE3.2* repeats. Inverted and direct repeats, mismatches, and gaps are indicated as in Fig. 1. Sizes of regions of the repeats not shown are indicated. The *asterisk* marks the target site of insertion of the *AthE1.4* element in the related *MAthE3.2* repeat in accession number Z97335 (Fig. 8). Positions in GenBank accession numbers for nucleotides shown are as follows: AB005233, 36,998–37,353; AF076275, 13,720–14,075; AC002339, 36,060–37,292 (reverse complement); and AC004481, 27,807–29,108.

ies in the available database. The *MAthE3.2* repeats are longer elements (approximately 1.2 kbp) defined by 70-bp terminal inverted repeats similar to the *MAthE3.1* repeats (Fig. 3B). They are also flanked by 8-bp direct repeats and are relatively infrequent (eight copies in the database). The two *MAthE3.2* elements shown in Fig. 3B are 93% conserved.

A Retroposon-like Repetitive Element

In addition to the MITE-like repetitive DNAs described above, another family, *RAthE1*, with many of the structural features of a nonviral retrotransposable element was identified. An alignment of *RAthE1* elements is shown in Fig. 4. These elements are approximately 170 bp in length, have well-defined ends, terminate in a 3′ oligo(A) tail, and are flanked by 8- to 11-bp direct repeats. These structural features are similar to short interspersed nuclear elements (SINES), retroelements common to eukaryotic genomes (Weiner et al. 1986). *RAthE1* elements are relatively common in the *Arabidopsis* genome; approximately 40 copies can be identified in the available database using sequences shown in Fig. 4.

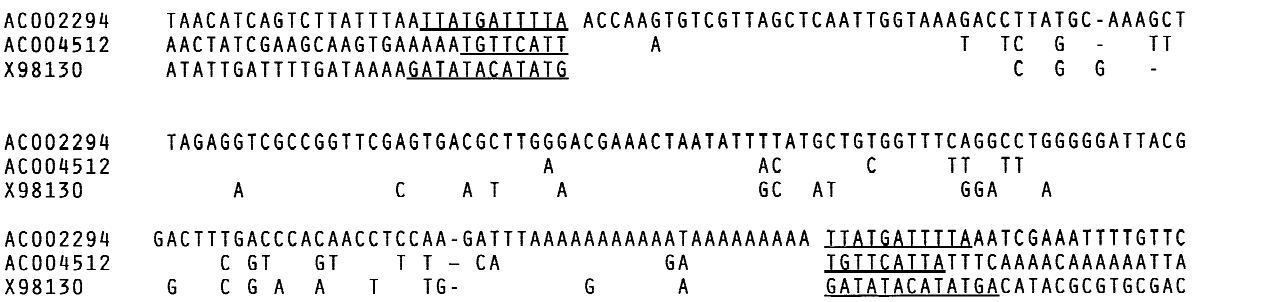

Fig. 4. Alignment of *RAthE1* repeats from *Arabidopsis.* Direct repeats, mismatches, and gaps are indicated as in Fig. 1. Positions in GenBank accession numbers for nucleotides shown are as follows: AC002294, 60,260–60,486 (reverse complement); AC004512, 83,436–83,660; and X98130, 33,537–33,763 (reverse complement).

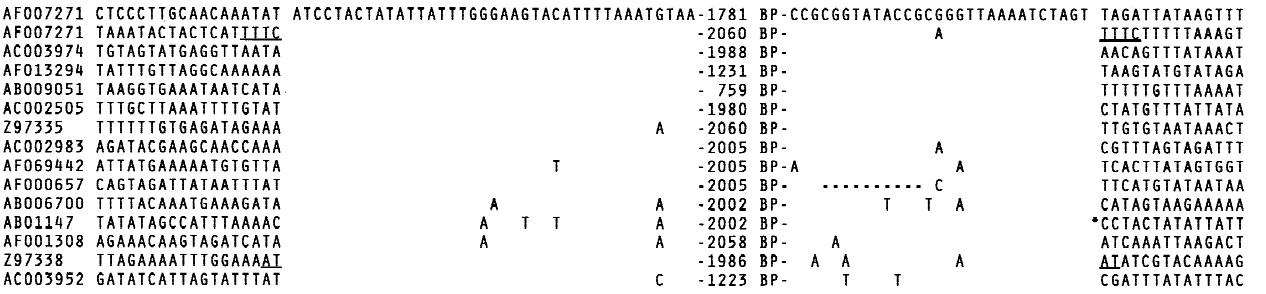

Fig. 5. Alignment of the 5′ and 3′ ends of *AthE1.4* repeats from *Arabidopsis.* Mismatches, gaps, and sizes of internal domains are indicated as in Fig. 1. Positions in GenBank accession numbers for nucleotides shown are as follows: AF007271, 62,750–64,360; AF007271, 29,725–31,885 (reverse complement); AC003974, 30,422–32,508; AF013294, 55,773–57,105; AB009051, 984–1842; AC002505, 70,766–72,844 (reverse complement); Z97335, 21,483–23,642; AC002983, 24,274–26,378; AF069442, 31,671–33,776 (reverse complement); AF000657, 54,195–56,291 (reverse complement); AB006700, 76,735–78,836; AB0011477, 30,903–33,004; AF001308, 89,925–92,084; Z97338, 93,035–95,121 (reverse complement); and AC003952, 9632–10,953. The *asterisk* in AB01147 indicates that the 3′ flanking sequence represents an adjacent *AthE1* element.

While a variety of retroelements has been described in the *Arabidopsis* genome (Chye et al. 1997; Konieczny et al. 1991; Pelissier et al. 1995; Wright et al. 1996; Wright and Voytas 1998), the *RAthE1* family members do not share sequence similarity with previously identified *Arabidopsis* retroposons.

The AthE1 *Family of Repetitive Elements*

The most common repetitive elements in *Arabidopsis* identified by this computational survey were the *AthE1* family members. The *AthE1* family is heterogeneous, however, highly similar members can be identified within the database. Figure 5 shows a partial alignment of three *AthE1* elements, in a subfamily denoted *AthE1.4.* The *AthE1.4* family is abundant in the *Arabidopsis* genome, and one these elements is contained within the previously described repeat ATR0053 (N.N. Dedhia and G.P. Copenhaver; AtRepBase, Cold Spring Harbor Laboratory). Full-length *AthE1.4* elements are approximately 2070 bp in length, and highly conserved members are easily identified. For example, alignment of elements from three BAC clones (GenBank accession numbers AF00657, AF069442, and AB006700) reveals 98% identity between them (data not shown). While the 5′ and 3′ ends of the *AthE1.4* elements are highly conserved (Fig. 5), the terminal sequences represent neither inverted or direct repeats. In addition to a number of full-length *AthE1.4* elements, severely deleted forms are also observed (Fig. 5). Finally, in contrast to the *MAthE* and *RAthE* elements (Figs. 1–4), these repetitive sequences are not, in general, flanked by direct repeats (Fig. 5).

By using the conserved 5′ and 3′ terminal sequences as database queries, a number of other *AthE1* subfamilies can be identified. While members of different subfamilies can contain little or no internal similarity, the terminal sequences which define the elements are conserved. For example, the partial sequences members of the family delineated *AthE1.1* are shown in Fig. 6. These elements are approximately 900 bp in length. *AthE1.1* elements located in three *Arabidopsis* BAC clones (GenBank accession numbers AB010695, AB012244, and AC001645) are approximately 95% conserved at the different loci. Similar to the *AthE1.4* family, examination of sequences flanking a number of *AthE1.1* family members (Fig. 6) reveals the absence of direct repeated domains. Comparison of the sequences of *AthE1.1* [AB0106975 (Fig. 6)] and *AthE1.4* [AF000657 (Fig. 5)] elements reveals that sequence similarity is limited to short domains in the 3′ terminal 100 bp (data not shown).

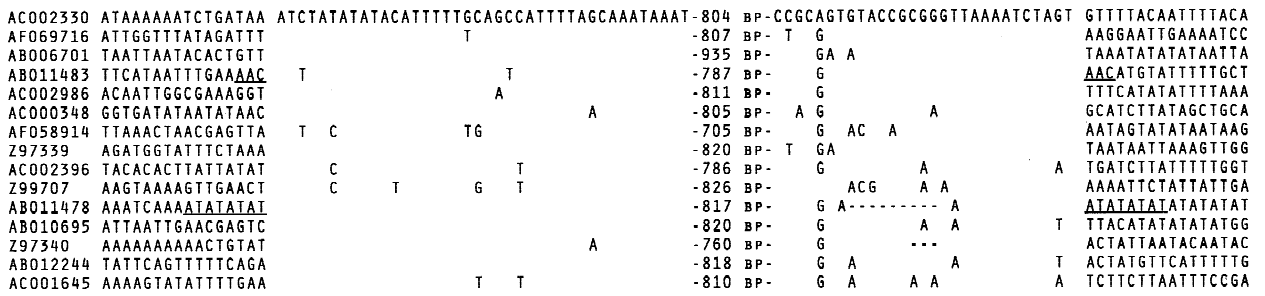

Fig. 6. Alignment of the 5′ and 3′ ends of *AthE1.1* repeats from *Arabidopsis.* Mismatches, gaps, and sizes of internal domains are indicated as in Fig. 1. Positions in GenBank accession numbers for nucleotides shown are as follows: AC002330, 11,493–12,397; AF069716, 11,055–11,962; AB006701, 68,266–69,301; AB011483, 34,125–35,011; AC002986, 84,458–85,369; AC000348, 25,603–26,510 (reverse complement); AF058914, 99,658–100,465 (reverse complement); Z97339, 67,706–68,626; AC002396, 102,052–102,938 (reverse complement); Z99707, 65,461–66,388 (reverse complement); AB011478, 45,070–45,979 (reverse complement); AB010695, 23,719–24,638; Z97340, 159,186–160,043; AB012244, 41,584–42,504 (reverse complement); and AC001645, 13,160–14,070.

Fig. 7. Consensus terminal regions of the *AthE1* subfamilies. Consensus sequences were determined within each subfamily. Alignment derived using AssemblyLIGN (Eastman Kodak Co.). Gaps and positions of repeat internal domains are indicated by *dashes.*

Figure 7 shows the consensus terminal sequences, and average insertion size, of nine subfamilies of *AthE1* repeats. Locations of representative elements for each subfamily (to assist in identification of additional subfamily members) and their approximate frequency in the available database are shown in Table 1. While within each subfamily highly conserved copies can be identified, internal sequence similarity between the subfamilies is limited. Of these subfamilies, the *AthE1.1, AthE1.2,* and *AthE1.3* repeats are the most closely related, with similarity at both the 5′ (100-bp) and the 3′ (150-bp) ends of the repeats. In addition, subfamilies *AthE1.7* and *AthE1.8* are related, with similarity extending 350 and 200 bp of the 5′ and 3′ ends of the elements, respectively. Similarity between other subfamilies is limited largely to the terminal 50-bp defining sequences (Fig. 7).

Similar to the data presented in Figs. 5 and 6, the absence of directly repeated sequences flanking the repeats is a general property of all the *AthE1* subfamilies. The lack of direct repeats flanking the *AthE1* elements indicates that they do not represent mobile DNAs capable of recombination into the genome by introducing the staggered cleavage sites in the target DNA, a feature common to both RNA- and DNA-based transposable elements (Berg and Howe 1989).

Three cases in which *AthE1.4* elements are inserted into other repetitive DNA sequences are shown in Fig. 8. In accession Z97335 the element is localized within a *MAthE3.2* repeat (Fig. 8A). When the sequences of three other *MAthE3.2* repeats are aligned to the flanking regions of the *AthE1.4* element, a high degree of similarity is observed. The alignment in Fig. 8A suggests that the recombination event resulting in the *AthE1.4* insertion in Z97335 results in a short (2-bp) deletion, rather than duplication, of the target sequence.

The *AthE1.4* element in accession AC004255 is localized within a 460-bp repeat found in two other *Arabidopsis* entries (Fig. 8B). Figure 8C shows an example in which an *AthE1.4* element is inserted into half of a direct repeat within the same locus (accession number

Table 1. Location of representative repeats, size, and frequency of *AthE1* subfamilies[a]

	Locus	Repeat position	Size (bp)	Frequency (No. copies)
AthE1.1	AC002396	18,201–19,075	880	35
Athe1.2	AB011485	38,033–38,591	550	37
AthE1.3	Z97340	11,3101–11,3681	590	17
AthE1.4	AB011477	30,921–32,990	2,100	55
Athe1.5	AF058826	39,149–41,380	2,200	12
AthE1.6	AB005236	24,719–25,733	1,000	23
AthE1.7	AF058914	34,794–35,974	1,200	7
AthE1.8	AC002396	56,480–57,380	900	7
AthE1.9	AB010694	39,460–40,189	750	6

[a] Representative repeats for each subfamily were identified by alignment of collected subfamily members using MacVector. Frequency indicates the number of copies identified in the available *Arabidopsis* database.

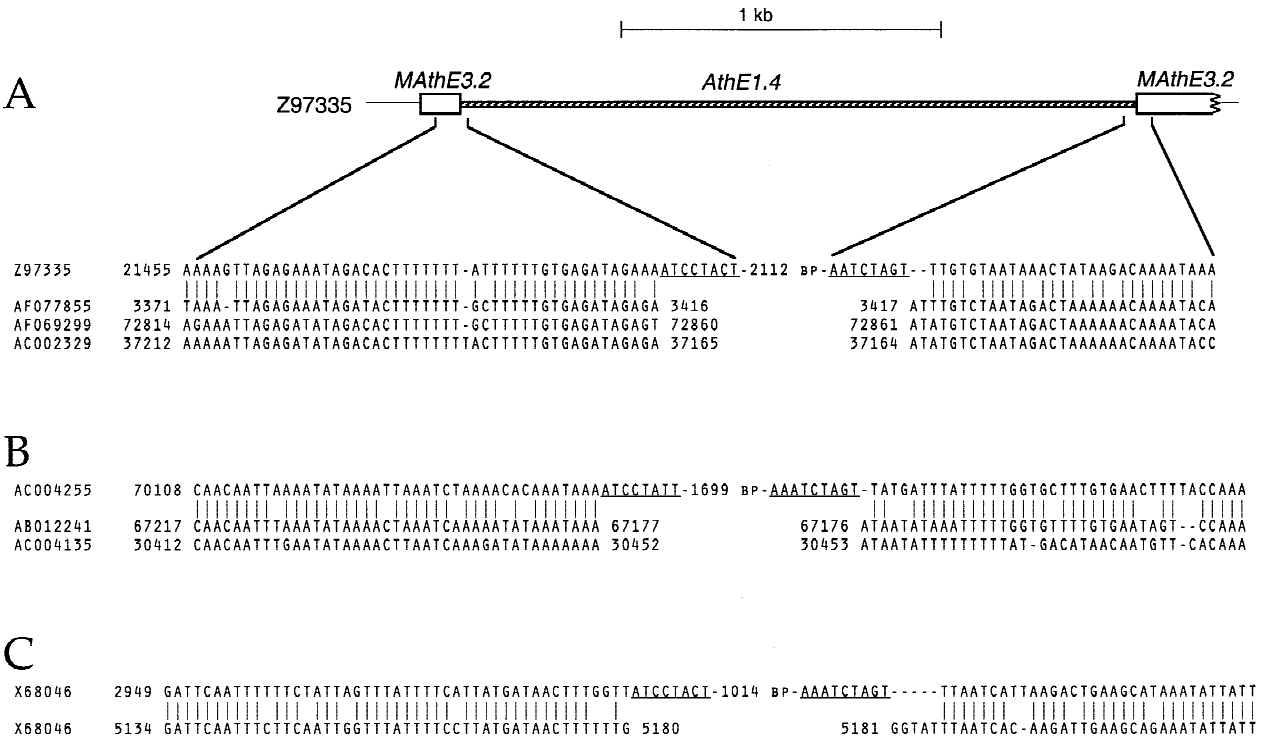

Fig. 8. Alignment of sequences flanking *AthE1.4* repeats with putative recombination target sequences. *AthE1.4* terminal sequences are *underlined;* gaps are indicated by *dashes. Horizontal lines* indicate matches between sequences flanking the repeats and at least one of the putative target sequences. Positions indicate positions of GenBank accession numbers.

X68046). Similar to the Z97335 element, no evidence for target site duplication upon recombinational insertion of the *AthE1.4* elements is observed. Once again, the data suggest that the recombination events associated with the insertion of the elements in AC004255 and X68046 result in short deletions of the target sequence (1 and 5 bp, respectively). The possibility exists that the five loci lacking the *AthE1.4* elements shown in Fig. 8 represent empty recombination sites following removal of the repeat. However, both the alignments shown in Fig. 8 and the data in Figs. 5 and 6 are consistent with recombination by an manner independent of the introduction of a staggered cut in the target sequence.

Frequency and Localization of AthE1 *Elements in the* Arabidopsis *Genome*

As indicated in Table 1, *AthE1* family members are common features in the *Arabidopsis* genome. The repeats are found on all chromosomes. On average, *AthE1* repeats are observed approximately every 100 kb of the genome.

Discussion

We describe a computational method for the identification of repetitive sequence domains in the *Arabidopsis* genome. The failure of our previous terminal inverted repeat-dependent algorithms (Oosumi et al. 1995a,b, 1996) to identify repetitive sequences in this organism may be a reflection of both the low abundance of repetitive DNA (Pruitt and Meyerowitz 1986) and the types of elements which are most common in this organism. While the computational survey described here revealed the expected MITE- and retroposon-like repeats (Wessler et al. 1995) (Figs. 1–4), the most abundant family of repeated sequences had features similar to neither.

The *AthE1* repeats are defined by conserved 5′ and 3′ terminal sequences which represent neither inverted nor direct repeats (Fig. 7). Many of the *AthE1* subfamilies share essentially no internal sequence similarity. For example, the *AthE1.4* and *AthE1.5* repeats are of similar size and their consensus 3′ terminal repeats are 90% identical (Fig. 7). However, when the full-length *AthE1.4* and *AthE1.5* elements listed in Table 1 are compared, similarity (>70% identity) is limited to a single 50-bp internal domain (data not shown). The presence of unrelated repetitive DNA elements flanked by similar terminal sequences is commonly observed in nonautonomous transposons, for example, *Ac/Ds* (Federoff 1989)- and *Mu* (Talbert et al. 1989)-based elements.

The most striking structural difference between the *AthE1* repeats and RNA- or DNA-based mobile elements is the general absence of direct repeats flanking the elements (Figs. 5 and 6). The direct repeats flanking mobile DNAs reflect a duplication caused by staggered cleavage of the target DNA sequence during recombination. These flanking direct repeats are an easily identifiable feature

of the mobile elements (Figs. 1–4), and their absence in flanking *AthE1* elements suggests a different mechanism of recombination. The data presented in Fig. 8 suggests that recombination events involving *AthE1* elements result in small deletions at the point of insertion in the target. These deletions, and the absence of flanking direct repeats, are consistent with *AthE1* insertion via illegitimate, similar to T-DNA integration events (Bundock and Hooykaas 1996; Gheysen et al. 1991; Gorbunova and Levy 1997; Mayerhofer et al. 1991).

The primary motivation for identification of repetitive DNA families in the *Arabidopsis* is the characterization of the evolutionary architecture of the genome. The results of the computational survey described here allow assignment of the evolutionary source of sequences comprising approximately 1% of the *Arabidopsis* genome. Of greater significance, these results allow definition of specific sequence domains within the *Arabidopsis* genome with potential to serve as sources of genetic variation in the evolutionary process. Finally, as indicated by the localization of a *MathE2*-like element in the tomato polyphenol oxidase A gene, identification of repetitive domains in *Arabidopsis* has the potential to facilitate characterization of the molecular architecture of other plant genomes.

Acknowledgments. The authors wish to express their gratitude to Benjamin Garlick (GigaPixel, Corp.) for expert technical assistance. References to a company and/or product by the USDA is only for purposes of information and does not imply approval or recommendation of the product to the exclusion of others that may also be suitable.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Berg, DE, Howe MM (eds) (1989) Mobile DNA. American Society for Microbiology, Washington, DC

Bevan M, Bancroft I, Bent E, et al. (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature 391:485–488

Britten RJ (1997) Mobile elements inserted in the distant past have taken on important functions. Gene 205:177–182

Bundock P, Hooykaas PJ (1996) Integration of Agrobacterium tumefaciens T-DNA in the Saccharomyces cerevisiae genome by illegitimate recombination. Proc Natl Acad Sci USA 93:15272–15275

Bureau TE, Wessler SR (1992) Tourist: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294

Bureau TE, Wessler SR (1994) Stowaway: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907–916

Bureau TE, Ronald PC, Wessler SR (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci USA 93:8524–8529

Chye ML, Cheung KY, Xu J (1997) Characterization of TSCL, a nonviral retroposon from Arabidopsis thaliana. Plant Mol Biol 35: 893–903

Fedoroff N (1989) Maize transposable elements. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, p 375

Frank MJ, Liu D, Tsay YF, Ustach C, Crawford NM (1997) Tag1 is an autonomous transposable element that shows somatic excision in both Arabidopsis and tobacco. Plant Cell 9:1745–1756

Gheysen G, Villarroel R, Van Montagu M (1991) Illegitimate recombination in plants: A model for T-DNA integration. Genes Dev 5:287–297

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res 25:4650–4657

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991) A superfamily of Arabidopsis thaliana retrotransposons. Genetics 127: 801–809

Lonnig WE, Saedler H (1997) Plant transposons: Contributors to evolution? Gene 205:245–253

Martinez-Zapater JM, Estelle MA, Somerville CC (1986) A highly repeated DNA sequence in Arabidopsis thaliana. Mol Gen Genet 204:417–423

Mayerhofer R, Koncz-Kalman Z, Nawrath C, et al. (1991) T-DNA integration: A mode of illegitimate recombination in plants. EMBO J 10:697–704

Newman SM, Eannetta NT, Yu H, et al. (1993) Organisation of the tomato polyphenol oxidase gene family. Plant Mol Biol 21:1035–1051

Ohta M, Sugita M, Sugiura M (1995) Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in Arabidopsis thaliana: Presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. Plant Mol Biol 27: 529–539

Oosumi T, Belknap WR (1997) Characterization of the *Sol3* family of nonautonomous transposable elements in tomato and potato. J Mol Evol 45:137–144

Oosumi T, Belknap WR, Garlick B (1995a) Mariner transposons in humans [letter]. Nature 378:672

Oosumi T, Garlick B, Belknap WR (1995b) Identification and characterization of putative transposable DNA elements in solanaceous plants and Caenorhabditis elegans. Proc Natl Acad Sci USA, 92: 8886–8890

Oosumi T, Garlick B, Belknap WR (1996) Identification of putative nonautonomous transposable elements associated with several transposon families in Caenorhabditis elegans. J Mol Evol 43:11–18

Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G (1995) Athila, a new retroelement from Arabidopsis thaliana. Plant Mol Biol 29:441–452

Pruitt RE, Meyerowitz EM (1986) Characterization of the genome of Arabidopsis thaliana. J Mol Biol 187:169–183

Richards EJ, Goodman HM, Ausubel FM (1991) The centromere region of Arabidopsis thaliana chromosome 1 contains telomere-similar sequences. Nucleic Acids Res 19:3351–3357

SanMiguel P, Tikhonov A, Jin YK, et al. (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Schmidt R, West J, Love K, et al. (1995) Physical map and organization of Arabidopsis thaliana chromosome 4. Science 270:480–483

Simoens CR, Gielen J, Van Montagu M, Inze D (1988) Characterization of highly repetitive sequences of Arabidopsis thaliana. Nucleic Acids Res 16:6753–6766

Talbert LE, Patterson GI, Chandler VL (1989) Mu transposable elements are structurally diverse and distributed throughout the genus Zea. J Mol Evol 29:28–39

Thompson H, Schmidt R, Brandes A, Heslop-Harrison JS, Dean C (1996a) A novel repetitive sequence associated with the centro-

meric regions of Arabidopsis thaliana chromosomes. Mol Gen Genet 253:247–252

Thompson HL, Schmidt R, Dean C (1996b) Analysis of the occurrence and nature of repeated DNA in an 850 kb region of Arabidopsis thaliana chromosome 4. Plant Mol Biol 32:553–7

Thompson HL, Schmidt R, Dean C (1996c) Identification and distribution of seven classes of middle-repetitive DNA in the Arabidopsis thaliana genome. Nucleic Acids Res 24:3017–3022

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Wessler SR (1996) Turned on by stress. Plant retrotransposons. Curr Biol 6:959–961

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Wright DA, Voytas DF (1998) Potential retroviruses in plants. Tat1 is related to a group of *Arabidopsis thaliana* ty3/gypsy retrotransposons that encode envelope- like proteins. Genetics 149:703–715

Wright DA, Ke N, Smalle J, Hauge BM, Goodman HM, Voytas DF (1996) Multiple non-LTR retrotransposons in the genome of Arabidopsis thaliana. Genetics 142:569–578