Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 8886–8890, September 1995
Genetics

Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*

TERUKO OOSUMI*, BENJAMIN GARLICK†, AND WILLIAM R. BELKNAP*‡

*U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, CA 94710; and †Silicon Graphics, Mountain View, CA 94039

Communicated by Robert Haselkorn, The University of Chicago, Chicago, IL, June 9, 1995 (received for review February 15, 1995)

ABSTRACT **Several families of putative transposable elements (TrEs) in both solanaceous plants and *Caenorhabditis elegans* have been identified by screening the DNA data base for inverted repeated domains present in multiple copies in the genome. The elements are localized within intron and flanking regions of many genes. These elements consist of two inverted repeats flanking sequences ranging from 5 bp to >500 bp. Identification of multiple elements in which sequence conservation includes both the flanking and internal regions implies that these TrEs are capable of duplicative transposition. Two of the elements were identified in promoter regions of the tomato (*Lycopersicon esculentum*) polygalacturonase and potato (*Solanum tuberosum*) *Win1* genes. The element in the polygalacturonase promoter spans a known regulatory region. In both cases, ancestral DNA sequences, which represent potential recombination target sequences prior to insertion of the elements, have been cloned from related species. The sequences of the inverted repeated domains in plants and *C. elegans* show a high degree of phylogenetic conservation. While frequency of the different elements is variable, some are present in very high copy number. A member of a single *C. elegans* TrE family is observed approximately once every 20 kb in the genome. The abundance of the described TrEs suggests utility in the genomic analysis of these and related organisms.**

A variety of transposable elements (TrEs) have been described in eukaryotic species. While in many cases the insertions of a TrE into the genome results in a deleterious phenotype, these elements have also been demonstrated to constitute integral components of a number of genes. In addition to introducing coding sequence (1), introns (2), and polyadenylylation signals (3), endogenous TrE sequences have been identified as providing important nuclear protein binding domains involved in regulation of eukaryotic genes (3–10). These TrEs provide a potential molecular mechanism for "enhancer shuffling," which allows the simultaneous introduction of multiple regulatory sequences and the addition of protein binding domains in an appropriate context (4, 6, 11).

In many cases, the TrE sequences that represent functional components of genes are defined by inverted repeated DNA sequences and are present at multiple genomic locations (3–5). To characterize more thoroughly the potential roles of TrEs in eukaryotic gene assembly, a systematic search of the GenBank nucleic acid data base for inverted repeated sequences was carried out. Several putative TrEs identified in solanaceous plants and *Caenorhabditis elegans* are presented here. These elements, which range in size from ≈100 bp to 1 kb, are not in general similar to previously identified elements and are only found in introns and flanking regions of genes. Additional elements were also found by probing the data base with the conserved inverted repeated domain of Tc2, a previously identified *C. elegans* transposon (12). This suggests that the TrEs described here may be nonautonomous derivatives of autonomous TrEs.

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.

MATERIALS AND METHODS

Cloning of Transposable Elements and Recombination Target Sequences. A homolog of the putative TrE localized in the tomato polygalacturonase (Tom*PG*) gene promoter, designated potato (Pot) *ten1*, was cloned from a potato λFIXII genomic library as described (13). The probe for the isolation of Pot*ten1* was PCR-generated by using a single oligonucleotide primer within the inverted repeat of the Tom*PG* element.

Target sequences for recombination events in the potato *Win1* (Pot*Win1*) (14) and Tom*PG* (15) promoters and the Pot*ten1* element were isolated by PCR with oligonucleotide primers flanking the putative TrEs. DNA from a selection of solanaceous species was screened for specific amplification of fragments with sizes consistent with target sequences prior to TrE insertion. Genomic DNA from *Solanum brevidans, Solanum chacoense, Solanum demissum*, and *Solanum etuberosum* was a kind gift from D. J. Hannapel (Iowa State University). For PCR primer design, sequences flanking the putative TrEs were analyzed by using RIGHTPRIMER (BioDisk, San Francisco), and primers 5′ and 3′ to the recombination events were selected. PCR products of appropriate sizes were amplified from the *S. chacoense* genomic DNA template for all three elements. Similar results were not obtained by using any of the other wild *Solanum* species. The *S. chacoense* reaction products were purified using a QIAquickspin PCR purification kit (Qiagen, Chatsworth, CA) and sequenced directly.

Computer Analysis. Inverted repeated sequences were found in the GenBank data base by using a specifically designed C language program incorporating an order(N) search algorithm. For each base (X) in the sequence, the program compared 750 of the bases on the 5′ side to the opposing bases on the 3′ side. To score the inverted repeats, complementing bases caused the counter to increase by one; mismatches caused the counter to decrease by one. During a string of mismatches, downward scoring was clamped at zero. The outward search position was saved where the highest counter value was recorded, marking the outer extremes of the strongest inverted repeat centered about base X. If the counter value was below a threshold value, the inverted repeat was ignored. The inverted repeat was also ignored if it was deemed to be the result of a self-complementing stutter. To disregard these stutters, a 4 × 4 transition frequency table was generated by incrementing an entry (L, R) for all L·R base pairs in the inverted repeat. If any of the 16 entries was overrepresented, i.e., (A, T) or (T, A), the inverted repeat was ignored. After testing of all X base pairs, the remaining strongest inverted

Abbreviations: TrE, transposable element; Mb, megabase(s); Tom (as a prefix), tomato; Pot (as a prefix), potato.

‡To whom reprint requests should be addressed at: U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710.

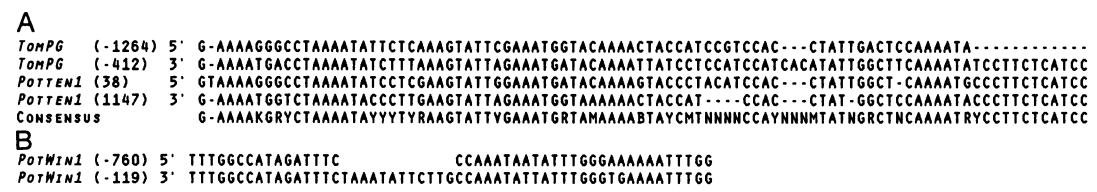

FIG. 1. Alignments of the inverted repeats from solanaceous plants. (*A*) Terminal regions of the *Sol3* class TrE inverted repeated domains of TrEs from potato (Pot*ten1*) and tomato (Tom*PG*). (*B*) Potato Pot*Win1* putative TrE inverted repeat. The 3′ inverted repeated sequences are shown as the reverse complement. Numbers indicate nucleotide positions relative to transcription start (Tom*PG* and Pot*Win1*) or to Fig. 2 (Pot*ten1*).

repeat candidates (if any) are reported to the user. This search operation is simple but computationally intensive. For this reason a Silicon Graphics (Mountain View, CA) super computer was used. The source code for the search program is available from the authors upon request. For analysis of the frequency of inverted repeated domains in *C. elegans* (see Table 2), a 1.1-megabase (Mb) contiguous assembly from chromosome III (16) was characterized by using the search program.

After identification of inverted repeated domains, combined data bases (GenBank and EMBL) were screened for related sequences by using the BLAST Network Service of the National Center for Biotechnology Information (17). Similar sequences were downloaded and evaluated for inverted repeated elements by using MACVECTOR (Eastman Kodak). Identification of similar regions in related TrE sequences was also carried out with MACVECTOR. ASSEMBLYLIGN (Eastman Kodak) was used in the derivation of consensus sequences.

RESULTS

Identification of Putative TrEs in Solanaceous Plants. A systematic search of GenBank for inverted repeated sequences was carried out and putative TrEs were screened for multiple copies within the genome. This screening took the form of both analysis of existing data banks and isolation of similar sequences from a genomic library.

Several inverted repeated elements from solanaceous plants were identified. The polygalacturonase gene in tomato Tom*PG* (locus TOMPGAAA, ref. 15) contains long inverted repeated sequences (≈250 bp) flanking 336 bp of internal sequence (Figs. 1*A* and 2*A*). The 3′ end of the 854-bp domain defined by the inverted repeats is located at position −412 with respect to the transcription start site (Fig. 2*A*). Inverted repeats are very common motifs in promoter elements. In many cases, key regulatory elements within the promoters are either flanked by or incorporated into inverted repeated sequences (4, 5). Montgomery *et al.* (18) identified a positive regulatory region in this control element at positions −806 to −433, a region included in the domain defined by the inverted repeats.

No sequences with similarity over the length of the putative TrE in Tom*PG* were identified by computational screening of currently available data base sequences. A potato genomic DNA library was, therefore, screened for sequences related to putative Tom*PG* TrE. Fig. 2*A* shows a sequence of the Tom*PG* TrE aligned to a similar element, Pot*ten1*, isolated by this procedure. The most significant deviation between the two elements is a 187-bp segment present in Pot*ten1* but absent in the Tom*PG* element. This insert is located immediately 5′ to the downstream arm of the repeat defining the tomato element.

Alignment of the terminal regions of the inverted repeated elements of the Tom*PG* and Pot*ten1* putative TrEs is shown in Fig. 1*A*. Two related elements within the Solanaceae were also identified in GenBank loci LERBCS2 (positions 2054–2276) and STPATP1 (positions 1686–2023). The inverted repeated domains of these elements share considerable similarity with

FIG. 2. Alignments of the putative TrEs in the tomato Tom*PG* and potato Pot*Win1* promoters with related elements from solanaceous plants. (*A*) Comparison of Tom*PG* sequence to potato Pot*ten1*. (*B*) Comparison of Pot*Win1* to the reverse complement of tomato locus TOMAPP1A. Only the mismatched base pairs are shown and gaps are indicated by dashes. Inverted repeated domains are indicated with italic type. Symbols in the Pot*ten1* sequence indicate insertions of 31 (*), 23 (+), and 187 (^) bp. The * in TOMAPP1A indicates the end of the entered sequence. Numbers indicate positions relative to transcription start. Initial nucleotide shown in TOMAPP1A is position 3834 in the GenBank entry.

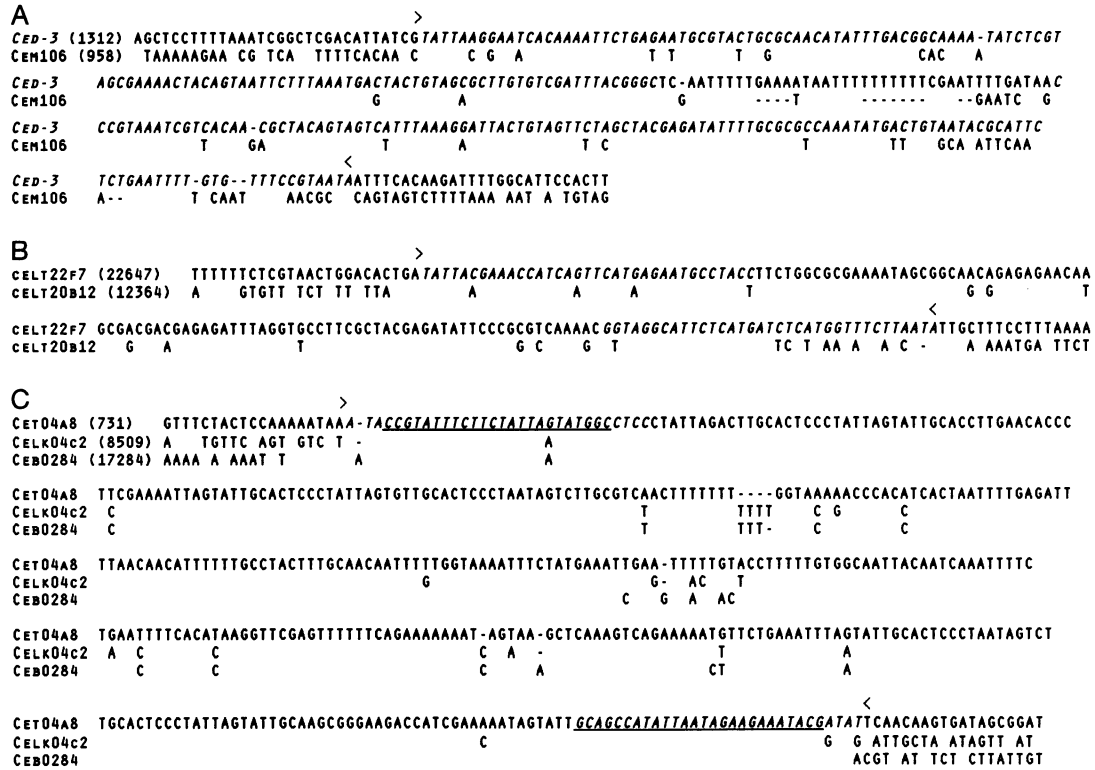

FIG. 3. Alignment of related putative TrEs from *C. elegans*. Related elements from the C*ele1* (*A*), C*ele6* (*B*), and C*eleTc2* (*C*) classes are aligned as in Fig. 2. The underlined sequences in *C* indicate the domain similar to the imperfect 24-bp inverted repeats that define the boundaries of the Tc2 transposon. Numbers indicate nucleotide positions in GenBank entries. Inverted repeated domains indicated as in Fig. 2.

those of Tom*PG* and Pot*en1*, suggesting that they belong to a single family, which we have designated *Sol3*.

The Pot*Win1* promoter region contains a domain defined by inverted repeated sequences at position −119 relative to the transcription start (Figs. 1*B* and 2*B*). This repeat flanks a 530-bp region proximal to the translation start that contains a "G-box" element at position −507 (14). The G-box element has been shown to be involved in the expression of a wide variety of wound-induced plant genes (19). Data base comparison using the putative TrE localized in the Pot*Win1* promoter revealed that a similar element is located 3′ to the acid phosphatase-1 gene in tomato (GenBank locus TOMAPP1A) (Fig. 2*B*). While only partial sequence of the element at the TOMAPP1A is available, similarity to the Pot*Win1* TrE region extends well into the internal sequence defined by the inverted repeated domain in the Pot*Win1* promoter.

Computational Analysis of the *C. elegans* Genome. The large quantity of sequence available in the *C. elegans* data base allowed identification of a number of putative TrEs in this organism. The sequence of an inverted repeat element in the promoter of the *C. elegans* cell death protein (*ced-3*) (GenBank locus CELCED3A) (20) is shown in Fig. 3*A*. This structure is defined by a 125-bp inverted repeat and is located proximal to the transcription start site (position −540). Comparison of the *ced-3* element to an element on chromosome III of *C. elegans* (GenBank locus CEM106) (16) is shown in Fig. 3*A*. Other TrEs in *C. elegans* with sequences similar to that of *ced-3* (for example, locus CET16H12, positions 3926–4230) are easily detected. As observed by Yuan *et al.* (20), inverted repeats similar to those associated with the TrEs in the *ced-3* locus are found in multiple locations in the *C. elegans* genome. Fig. 4 shows the highly conserved nature of the long repeated elements of the *ced-3*-related elements, which we have designated *Cele1*, at a number of genomic locations.

Five additional conserved inverted repeat domains, which are present at high copy number in the *C. elegans* genome, are listed in Table 1. The number of inverted pairs shown in Table 1 indicates only those putative TrEs in which both arms were identified as high scoring matches. In almost all cases (>90%), complementary sequences can be identified under lower stringency. However, failure to identify an inverted pair of a high scoring match can also occur when the putative TrE sequence exists at the boundary of a reported sequence. For example, GenBank locus CELF42H10, contains a sequence complementary to the C*ele2* domain (Table 1) at positions 1–57. As observed with the *ced-3* element, essentially identical elements, including both the inverted repeated domains and internal sequence, are easily identified in the *C. elegans* data base. Two examples are presented in Fig. 3 *B* and *C*. Elements

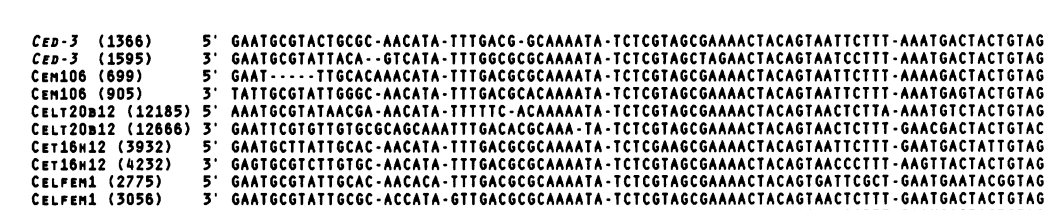

FIG. 4. Alignments of C*ele1* class inverted repeated domains of putative TrEs for *C. elegans*. Alignments are as in Fig. 1. Numbers indicate nucleotide positions in GenBank entries.

Table 1. Consensus sequences of putative *C. elegans* transposable element inverted repeats

Element	Inverted repeat consensus sequence	High scoring matches	Inverted pairs
Cele1	CAAAATATCTCGTAGCGAAAACTACAGTAAYTCTTT	99	42
Cele2	TACCHGGTCTCGACACGACANNTTTNTVTTNAATNNRAANNGDTGTGH-GCCTTTAAAGADTACTGTANTTTNAAANTTTNGTTDCTGCNNAATTTT	>400	>170
Cele4	TGGGTCTCGTTAGGTATTHGVNGGCAAAAHYVTVAATT	71	28
Cele5	GGTCTCGAAACGAYYGAAAYTTYGHAGCTACCGTAHC	44	14
Cele6	TATTAMGRRAHCAHNARWTCATGAGAATGCBTA	22	8
Cele7	TAGTGHNAAANTATAGAAAATYAYTTGNTTTWTCTGAAAATAACDTH-DATTTBNNBTNGAAHHWTTNGARAAA	77	31

Consensus sequences of inverted repeated domains are derived and presented as in Fig. 2. The query sequences submitted to the BLAST Network Server were derived from the above consensus sequences but contained only A, C, G, T, or N. All searches used the default matrix. High scoring matches are defined as having a Poisson *P* value of <0.001. In all cases, returned sequences with *P(N)* less than this value were derived from *C. elegans*. Inverse pairs indicates two high scoring elements in inverse orientation within 1.5 kb.

essentially identical to the *Cele6* putative TrEs (Table 1 and Fig. 3*B*) are found in other locations in *C. elegans* (CELZK637 position 26930 and CELF37C12G position 31058). It is of interest to note that when a highly conserved 27-bp subsequence internal to the *Cele2* repeat (5′-TACCCGGTCTC-GACGTGACAAATTTTT-3′) is used as a query sequence for data base searching, all 75 high scoring matches returned are derived from *C. elegans*.

To characterize the frequency of the putative TrEs in the *C. elegans* genome, a 1.2-Mb contiguous assembly from chromosome III (16) (position 1 GenBank locus CELZK112 to position 20,000 locus CELF44E2) was searched for inverted repeated sequences (Table 2). Analysis returned a total of 55 discrete domains of <1.5 kb defined by inverted repeated sequences with scores >20. The length of the complementing inverted repeated sequences in this data set varied from 23 to 406 bp. The average length of the inverted repeats in this data set was ≈65 bp. Given the scoring algorithm used, the percent mismatch in each inverted repeat varies inversely with length of the repeat. As shown in Fig. 4, the sequences of some of the putative TrEs are highly conserved. In repeats >40 bp long, ≈15% mismatching is observed (data not shown).

As indicated in Table 2, 21 of the high scoring inverted repeated sequences identified in the 1.2-Mb assembly can be assigned to putative TrE families by sequence similarity to either the elements in Table 1 or *C. elegans* transposons Tc2 (12) and Tc5 (21). Additional putative TrEs, with scores <20, were also found within the cosmid sequences making up the 1.2-Mb assembly (Table 2). A total of 36 *Cele2* elements are found in this contig, with an average length of 345 bp, indicating that this putative TrE accounts for ≈1% of the total sequence.

Table 2. Inverted repeated domains in 1.2 Mb of contiguous *C. elegans* sequence

Total no. inverted repeats with scores ≥20	55
No. elements with scores ≥20 similar to	
Cele1	2
Cele2	14
Cele4	2
Tc2	2
Tc5	1
No. elements with scores <20* similar to	
Cele1	4
Cele2	22
Cele4	4

A contiguous 1.2-Mb portion of *C. elegans* chromosome III (16) was analyzed for domains of <1.5 kb defined by inverted repeated sequences. Discrete domains were classified by sequence similarity of the inverted repeats to TrEs of the same class (Table 1) or to known *C. elegans* transposons (Tc2 and Tc5).

*Elements identified within GenBank submitted cosmid sequences constituting parts of the 1.2-Mb contiguous sequence as described in Table 1.

Identification of Target Sequences for TrE Recombination. Four potential target sequences for recombination events involving TrEs in solanaceous plants have been identified. A target sequence for recombination of the Tom*PG* TrE was isolated by PCR from another solanaceous species (*S. chacoense*), shown in Fig. 5*A*. Recombination target sequences for both the Pot*ten1* and Pot*Win1* elements (Fig. 5 *B* and *D*) were similarly isolated. For the putative TrE in the first intron of the patatin pseudogene STPATP1, comparison with the same location of an active patatin gene (locus POTPATA) revealed the absence of a TrE at this locus (Fig. 5*C*). The degradation in sequence similarity between the target sequences and those flanking the Tom*PG* element do not permit an accurate characterization of recombination events (Fig. 5*A*). In contrast, in the Pot*ten1* and STPATP1 target sequences from solanaceae (Fig. 5 *B* and *C*) sequence alignments are consistent with duplication of 4 to 5 bp of target sequence upon recombination. For the Pot*Win1* TrE, the precise definition of the borders of the element are unclear (Fig. 2*B*). The alignment presented in Fig. 5*D* could reflect a 2-bp duplication upon TrE isolation or an imprecise excision event.

Similar to the TrEs (3), nested sets of the *C. elegans* elements are also observed. An example of this type of structure (GenBank locus, CELT20B12) is shown in Fig. 5*E*, in which a *Cele1* element is interrupted by an element with *Cele6* repeats. The sequence of the interrupting element, found proximal to the 3′ *Cele1* inverted repeat is shown in Fig. 3*B*. In Fig. 5*E*, the sequence of the interrupted *Cele1* (CELT20B12) element is aligned to a related element that lacks the *Cele6* element (see also CELR05D3 position 19,735 for element related to CEC07A9). These data are also consistent with integration of the *Cele6* element (Fig. 3*B*) into a *Cele1* element related to those found in GenBank loci CEC07A9 and CELR05D3.

TrEs Related to the *C. elegans* Transposon Tc2. The *Ac/Ds* TrE family in maize includes both autonomous *Ac* and non-autonomous *Ds* elements. While these elements are defined by the same 11-bp inverted repeated sequences, many *Ds* elements vary considerably in both size and internal sequence (22, 23). Similarly, when the 24-bp inverted repeat sequence from Tc2 (12) was used as a data base query, in addition to the Tc2 transposon, a heterogeneous family of putative TrEs was identified. These elements ranged in size from 99 bp (GenBank locus CELD27H5, positions 13,754–13,851) to 500 bp (Fig. 3*C*). Two closely related elements defined by Tc2 inverted repeats are compared in Fig. 3*C* (a similar element is located at position 17,284 of locus CEB0284). While the internal sequences of the TrEs in this figure are highly conserved, they have very limited similarity to the autonomous Tc2 transposon. This family of elements thus shows a striking similarity with the

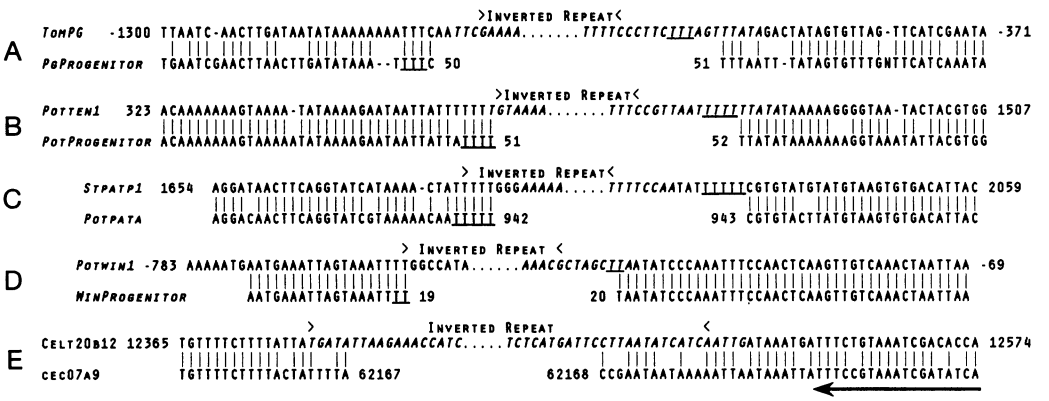

FIG. 5. Alignment of sequences flanking TrEs with putative recombination target sequences. Underlined sequences reflect potential duplicated target sequences. Boundaries of the inverted repeats are indicated by > (5′) and < (3′). Bases in italic type indicate regions of similarity between related TrEs: *A, B,* and *D*, determined from sequences in Fig. 2; *C*, determined from alignment of related TrEs in STPATP1 and LERBCS2; *E*, determined from sequence alignment in Fig. 3*B* and closely related elements described in text. Sequence numbering for Tom*PG* and Pot*Win1* indicate position relative to transcription start; other positions indicate nucleotide positions in GenBank entries (STPATP1, POTPATA, CELT20B12, CEC07A9) or PCR fragments (PgProgenitor, PotProgenitor, WinProgenitor). The bold arrow (*E*) indicates the boundary of the 3′ C*ele1* inverted repeat.

Ac/Ds family and suggests a potential source of the other putative TrEs described here.

DISCUSSION

We describe a method for identification of putative TrEs. Several lines of evidence suggest that the elements described here represent TrEs. (*i*) Elements in which both the inverted repeated domains and internal sequences are highly conserved can be found in multiple locations in the genome (24). (*ii*) Sequence similarity between loci containing related elements ends at, or near, the ends of the inverted repeats. (*iii*) Sequences that represent potential target sites for recombination can be identified.

The elements described here are localized to regions 5′ and 3′ to identified genes and within introns. A number of cases have now been reported in which key nuclear protein binding domains have been introduced into promoter elements via endogenous TrEs (3, 6–10). The putative TrE in the Tom*PG* promoter described here introduced three nuclear protein binding domains (Fig. 2*A*). While the nuclear binding domains of the Pot*Win1* and *ced-3* genes (Fig. 3*A*) have not yet been characterized, the proximity of the elements to the transcription start site in each case and the presence of a G-box (21) in the Pot*Win1* element suggest the potential for a similar contribution of regulatory elements.

In addition to defining, at least in part, the evolutionary architecture of a given locus, the ability to define accurately the regions of a sequence that are composed of TrEs has important implications in the analysis component of large-scale DNA sequencing programs. This is clearly true in *C. elegans*, where a single family of TrEs identified here (C*ele2*; Table 1) accounts for ≈1% of the genome. As the families shown in Table 1 represent only a partial list of *C. elegans* TrEs, it is clear that a complete compilation of these elements would serve as a valuable tool in characterizing a significant percentage of this organisms genome.

We express our gratitude to S. Wessler, T. Bureau, J. Collins, W. D. Park, and J. K. Belknap for helpful discussion and to D. Rockhold for expert technical assistance.

1. Banki, K., Halladay, D. & Perl, A. (1994) *J. Biol. Chem.* **269,** 2847–2851.
2. Wessler, S. R. (1989) *Gene* **82,** 127–133.
3. Bureau, T. E. & Wessler, S. R. (1994) *Plant Cell* **6,** 907–916.
4. Anderson, R., Britten, R. J. & Davidson, E. H. (1994) *Dev. Biol.* **163,** 11–18.
5. Dariavach, P., Williams, G. T., Campbell, K., Pettersson, S. & Neuberger, M. S. (1991) *Eur. J. Immunol.* **21,** 1499–1504.
6. White, S. E., Habera, L. F. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11792–11796.
7. Stavenhagen, J. B. & Robins, D. M. (1988) *Cell* **55,** 247–254.
8. Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C. & Meisler, M. H. (1992) *Genes Dev.* **6,** 1457–1465.
9. Banville, D. & Boie, Y. (1989) *J. Mol. Biol.* **207,** 481–490.
10. Chang-Yeh, A., Mold, D. E. & Huang, R. C. (1991) *Nucleic Acids Res.* **19,** 3667–3672.
11. Reue, K., Leff, T. & Breslow, J. L. (1988) *J. Biol. Chem.* **263,** 6857–6864.
12. Ruvolo, V., Hill, J. E. & Levitt, A. (1992) *DNA Cell. Biol.* **11,** 111–122.
13. Garbarino, J. E. & Belknap, W. R. (1994) *Plant Mol. Biol.* **24,** 119–127.
14. Stanford, A., Bevan, M. & Northcote, D. (1989) *Mol. Gen. Genet.* **215,** 200–208.
15. Bird, C. R., Smith, C. J. S., Ray, J. A., Moureau, P., Bevan, M. W., Bird, A. S., Hughes, S., Morries, P. C., Grierson, D. & Schuch, W. (1988) *Plant Mol. Biol.* **11,** 651–662.
16. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., *et al.* (1994) *Nature (London)* **368,** 32–38.
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
18. Montgomery, J., Pollard, V., Deikman, J. & Fishcer, R. L. (1993) *Plant Cell* **5,** 1049–1062.
19. Foster, R., Izawa, T. & Chua, N. H. (1994) *FASEB J.* **8,** 192–200.
20. Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. (1993) *Cell* **75,** 641–652.
21. Olsen, P., Andrews, S. & Collins, J. (1994) *Genetics* **137,** 771–781.
22. Federoff, N., Wessler, S. & Shure, M. (1983) *Cell* **35,** 235–242.
23. Sutton, W. D., Gerlach, W. L., Schwartz, D. & Peacock, W. J. (1984) *Science* **223,** 1265–1268.
24. Berg, D. E. & Howe, M. M. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington, DC).