

G. Astua-Monge · A. Lyznik · V. Jones
S.A. Mackenzie · C.E. Vallejos

Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases

Received: 5 December 2000 / Accepted: 25 April 2001

Abstract An insertion-sequence of prokaryotic origin was detected in a genomic clone obtained from a *Phaseolus vulgaris* bacterial artificial chromosome (BAC) library. This BAC clone, characterized as part of a contig constructed near a virus resistance gene, exhibited restriction fragment length polymorphism with an overlapping clone of the contig. Restriction analysis of DNA obtained from individual colonies of the stock culture indicated the presence of a mixed population of wild-type and insertional mutants. Sequence analysis of both members of the population revealed the presence of *IS10R*, an insertion-sequence from *Escherichia coli*. A BLAST search for *IS10*-like sequences detected unexpected homologies with a large number of eukaryotic sequences from *Homo sapiens*, *Arabidopsis thaliana*, *Drosophila melanogaster* and *Caenorhabditis elegans*. Southern analysis of a random sample of BAC clones failed to detect *IS10* in the BAC DNA. However, prolonged subculturing of a set of 15 clones resulted in transposition into the BAC DNA. Eventually, all cultures acquired a 2.3-kb fragment that hybridized strongly with *IS10*. Sequence analysis revealed the presence of a preferred site for transposition in the BAC vector. These results indicate that a large number, if not all, of the BAC libraries from different organisms are contaminated with *IS10R*. The source of this element has been identified as the DH10B strain of *E. coli* used as the host for BAC libraries.

Keywords *IS10* · Tn10 · BAC libraries

Communicated by M.A. Saghai Maroof

G. Astua-Monge · V. Jones · C.E. Vallejos (✉)
Department of Horticultural Sciences,
and Plant Molecular and Cellular Biology Graduate Program,
University of Florida, Gainesville, FL 32611-0690, USA

A. Lyznik · S.A. Mackenzie
Plant Science Initiative, Beadle Center for Genetics Research,
University of Nebraska, Lincoln, NE 68588-0665, USA

Introduction

Nucleic acid databases have become a powerful tool for gene discovery and comparative genomics. Unfortunately, a relatively high proportion of entries have been found to be contaminated with extraneous sequences. While some sources of contamination are endogenous, e.g., ribosomal RNA sequences (Gonzalez and Sylvester 1997), the most common sources are plasmid vector sequences, bacteriophage M13 (Miller et al. 1999; Seluja et al. 1999), and sequences from the cloning host (Binns 1993; Jha and Sarkar 1997). *Escherichia coli* insertion-sequences (IS) have also been found as contaminants of cloned eukaryotic sequences (Binns et al. 1986; Binns 1993).

IS1 and *IS421* were reported to be the most common IS elements contaminating database entries (Binns 1993). In fact, a recent homology search detected eight eukaryotic sequences contaminated with *IS1* and 13 with *IS421*. However, in recent years there has not been any new reports of additional IS sequences as contaminants of eukaryotic sequences. We report here the detection of a prokaryotic transposable element in the insert of a BAC clone of *Phaseolus vulgaris* (the common bean). This element is also present in several entries of common databases, including several from the human genome.

Materials and methods

Extraction of BAC DNA

BAC clones were obtained from a *P. vulgaris* BAC library (Vanhouten and Mackenzie 1999). *E. coli* carrying the BAC clones was grown in Luria-Bertani (LB) medium (Sambrook et al. 1989) containing 34 mg/ml of chloramphenicol. Cultures were grown for 16 h at 37°C with continuous agitation at 250 rpm. BAC DNA was extracted using a BAC DNA purification kit from Pinceton Separations, Incorporated (Adelphia, N.J.), or a modified alkaline-lysis method (Sambrook et al. 1989). In the latter method, bacterial cells (2 ml) were pelleted by centrifugation at 1500×g_{max} for 15 min, resuspended in 200 µl of GTE buffer (50 mM Glucose, 25 mM Tris.HCl pH 7.8, 10 mM Na EDTA) and lysed by the addi-

tion of 300 μ l of a solution containing 0.2 M NaOH/1% SDS at room temperature for no more than 5 min. The mixture was neutralized for 5 min on ice with 300 μ l of 5 M potassium acetate (pH 4.5). The lysate was centrifuged for 10 min at 21000 g_{max} and the resulting supernatant was transferred to a new tube for extraction with 1 vol of chloroform. After inverting the tube by hand 4 to 5 times, the phases were separated at 21000 g_{max} for 5 min. The upper phase was mixed with 1 vol of isopropanol and the DNA was pelleted at 21000 g_{max} for 30 min. The pellet was washed in 500 μ l of 100% ethanol, then dried and resuspended in 30 μ l of 0.1 \times TE buffer. In contrast to results obtained with the purification kit, alkaline lysis preparations tended to have some chromosomal DNA contamination.

Southern analysis

BAC DNA was digested with *EcoRI* or *EcoRV* according to the manufacturer's instructions (Life Technologies, Grand Island, N.Y.). DNA restriction fragments were resolved by agarose-gel electrophoresis (Sambrook et al. 1989) and transferred to a Biotodyne nylon membrane (Life Technologies, Grand Island, N.Y.) as described by Southern (1975) and modified by Chomczynsky (1992). The probes were labeled with ^{32}P using a Random Primed DNA Labeling kit (Boehringer Mannheim, Indianapolis, Ind.), or PCR (Plyler and Vallejos, 2000). Primers for PCR-amplification of *IS10R* were designed with Primer3 ([http://www_genome.wi.mit.edu/cgi_bin/primer/primer3.cgi](http://www.genome.wi.mit.edu/cgi_bin/primer/primer3.cgi)). These primers [(5'CCTC-ATAATTTCCCAAAG3') and (5'GCGAATGCAGATTGAAGAAAC3')] amplified a 351-bp product. Hybridization signals were visualized by autoradiography on X-OMAT film (Kodak).

DNA sequencing

Sequencing was carried out at the Interdisciplinary Center for Biotechnology Research (ICBR) Sequencing Core Lab (University of Florida, Gainesville, Fla.) and the Iowa State University DNA Sequence Facility (Ames, Iowa) using the Applied Biosystems model 373 or 377 Prism system (Applied Biosystems, Foster City, Calif.). The computer programs SeqAid II version 3.81 (Shareware by D. D. Rhoads and D. J. Roufa, Kansas State University) and BLAST 2.0 (Altschul et al. 1997) were used to analyze nucleotide sequence data and predicted protein products.

Results

Detection and identification of the contaminating transposable element

We are using a *P. vulgaris* (the common bean) BAC library (Vanhouten and Mackenzie 1999) to construct a contig around a virus resistance gene. For the purpose of this report we focus on two overlapping clones of the contig: BnBAC25 and BnBAC55. Restriction fragments containing the termini of BnBAC25 were isolated and cloned previously (Plyler and Vallejos 2000). BAC25L was identified as containing one of the termini of BnBAC25 and also as one of the leading ends of the contig. Southern hybridizations of this clone to genomic blots of *P. vulgaris* DNA digested with six restriction enzymes indicated that this clone represents a single-copy genomic sequence. BAC25L was used to screen the *P. vulgaris* BAC library and isolate BnBAC55. Overlap between these clones was inferred by analysis of restriction fragment profiles and by hybridization of Southern blots with ^{32}P -labeled BAC25L. Southern hybridization of *EcoRI*-digested BnBAC25 and BnBAC55 clones

Fig. 1A–D Southern-hybridization patterns of *EcoRI* digests of *P. vulgaris* BAC clones resulting from the insertion of the prokaryotic transposable element *IS10R*. The numbers on each lane in panels A and C represent the BnBAC clone designation, and those on panel B the colony number of BnBAC55. **A** A southern-blot probed with BAC25L showing that BnBAC 25 and BnBAC55 share a 2.8-kb *EcoRI* fragment, but BnBAC55 possesses an additional 4.1-kb cross-hybridizing fragment. **B** A southern-blot of individual BnBAC55 colonies showing the 2.8- or the 4.1-kb fragment that hybridizes to BAC25L. **C** A southern-blot of several BAC clones showing that only the mutant BnBAC55::*IS10R* has a 4.1-kb fragment that hybridizes to a probe derived from the *IS10R* sequence. **D** Graphic representation of the results depicted in the other panels

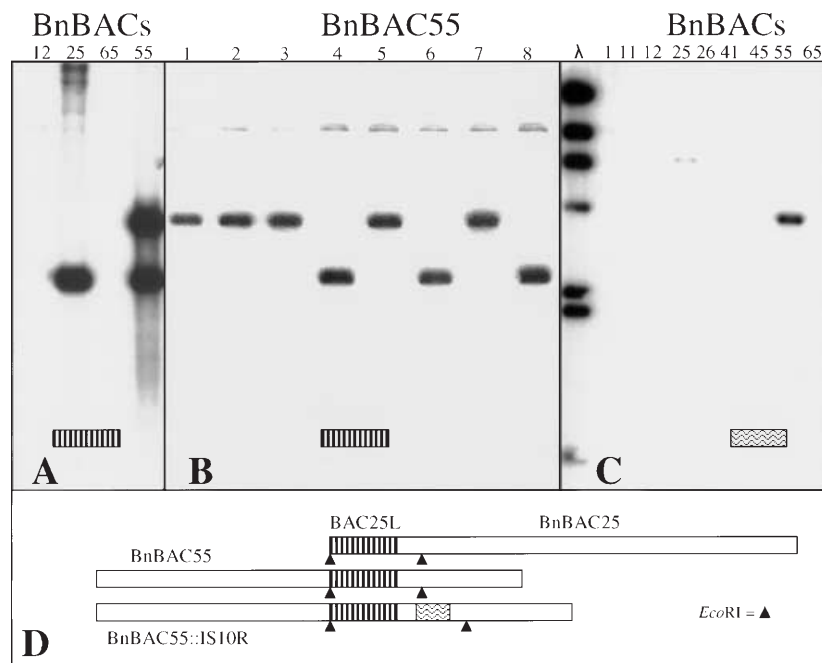


Table 1 Sequences with identity to the IS element found in a *P. vulgaris* BAC library. Scores and E-values were as determined by Gapped BLAST 2.0 (Altschul et al. 1997)

Database ID number	Sequence origin	Score	E value
gb J01829.1	Transposon Tn10 DNA, 3' end	2635	0.0
gb AC007396.5	<i>Arabidopsis thaliana</i> , BAC T4O12	2635	0.0
gb AC007887.8	<i>Arabidopsis thaliana</i> , BAC F15O4	2635	0.0
gb AC012463.3	<i>Arabidopsis thaliana</i> , BAC T2E6, Chromosome 1	2555	0.0
gb AC006650.1	<i>Caenorhabditis elegans</i> , H05LO3	2555	0.0
gb AF181652.1	<i>Drosophila melanogaster</i> , BAC LD23587	2635	0.0
gb AE003309.1	<i>Drosophila melanogaster</i> , genomic scaffold	2508	0.0
gb AC008583.4	<i>Homo sapiens</i> chromosome 16, CTC-566I14	2635	0.0
gb AC007226.2	<i>Homo sapiens</i> chromosome 16, RPCI-11 520B23	2635	0.0
gb AC005410.1	<i>Homo sapiens</i> chromosome 17, hRPK-1096 G 20	2635	0.0
gb AC008102.16	<i>Homo sapiens</i> chromosome 11, 11q13, BAC 41I21	2635	0.0
gb AC009247.11	<i>Homo sapiens</i> , 3q27, BAC RP11-410F19	2635	0.0
gb AC004617.1	<i>Homo sapiens</i> chromosome Y, 264M20	2635	0.0
gb AC007630.3	<i>Homo sapiens</i> , P1 clone P17739	2635	0.0
gb AC007566.1	<i>Homo sapiens</i> , clone RG010G	2635	0.0
gb AC006210.1	<i>Homo sapiens</i> , Xp22-150 BAC GSHB-309P15	2635	0.0
gb AC005552.1	<i>Homo sapiens</i> , chromosome 17, hRPK.212 E8	2635	0.0
gb AC005215.1	<i>Homo sapiens</i> , chromosome 5, and BAC 89K19	2635	0.0
emb AL136059.2	<i>Homo sapiens</i> , chromosome 5, BAC R-391B7	2635	0.0
emb AL137818.2	<i>Homo sapiens</i> , chromosome 14, BAC R-102C24	2635	0.0
dbj AP000025.1	<i>Homo sapiens</i> , genomic DNA from 21q11.1	2635	0.0
dbj AP000245.1	<i>Homo sapiens</i> , genomic DNA from 21q11.1	2635	0.0
dbj AP000128.1	<i>Homo sapiens</i> , genomic DNA from 21q11.1	2635	0.0
emb AL355112.1	<i>Homo sapiens</i> , chromosome 14, BAC R_159D23	2635	0.0
gb AC005317.1	<i>Homo sapiens</i> , 15q26.1, PAC pDJ457J11	2619	0.0
gb AC006305.1	<i>Homo sapiens</i> , chromosome 18, hRPK.456O19	2555	0.0
emb AL133246.2	<i>Homo sapiens</i> , region 2p21-2p22, BAC 852C13	2555	0.0
emb AL117609.1	<i>Homo sapiens</i> , cDNA DKFZp564O0463	2555	0.0
emb AL133224.2	<i>Homo sapiens</i> , chromosome 14, BAC R-16G24	2555	0.0
dbj AP000056.1	<i>Homo sapiens</i> , genomic DNA from 21q22.1	2555	0.0
dbj AP000330.1	<i>Homo sapiens</i> , genomic DNA from 21q22.1	2555	0.0
dbj AP000124.1	<i>Homo sapiens</i> , genomic DNA from 21q22.1	2555	0.0
gb AF199339.1	<i>Homo sapiens</i> , lens epithelium-derived GF	2547	0.0
gb AC006473.2	<i>Homo sapiens</i> , chromosome 17, hRPK-203M16	2508	0.0
dbj AK001627.1	<i>Homo sapiens</i> , cDNA FLJ107	2508	0.0
gb AC015600.6	<i>Homo sapiens</i> , clone RP11_45C5	2508	0.0
gb AC008496.4	<i>Homo sapiens</i> , chromosome 5 clone CTC_431G16	2508	0.0
emb AJ001004.1	<i>Picea abies</i> 8.7-kb chloroplast, Lambda Zap II clone	2454	0.0

probed with BAC25L revealed the presence of a 2.8-kb fragment in both clones. This is exactly what would be expected of overlapping clones because the *P. vulgaris* BAC library was constructed with partially digested *EcoRI* fragments.

When the hybridization described above was repeated a few months later with a new Southern blot of the BAC clones, both BnBAC25 and BnBAC55 displayed the same 2.8-kb restriction fragment as before, but BnBAC55 displayed an additional 4.1-kb fragment (Fig. 1A). Three explanations were considered for the new fragment in BnBAC55: (1) partial digestion, (2) a mutation at one of the *EcoRI* sites, or (3) an insertion event. The first explanation was discarded because inspection of the *EcoRI* restriction profile of BnBAC55 showed that the digestion had been carried out to completion. To discriminate between the latter two alternatives, BnBAC55 cells from the original freezer stock were plated on chloramphenicol plates and eight single colonies were selected for BAC DNA extraction and analysis. These DNAs were digested with *EcoRI*, blotted onto nylon membranes and hybridized with BAC25L.

Figure 1B shows that one group of colonies contains the 2.8-kb fragment while the other group contains the 4.1-kb fragment. To identify the nature of the shift in molecular weight, one colony from each group was selected for sequencing in the region encompassing BAC25L. Primers derived from the BAC25L sequence were used for this purpose and for sequencing the region adjacent to BAC25L in BnBAC25 as well.

Sequence comparisons among BnBAC25, the BnBAC55 wild-type and the BnBAC55 mutant revealed the presence of a 1329 bp insertion in the mutant. A sequence-homology search of the insertion-sequence with Gapped BLAST 2.0 (Altschul et al. 1997) indicated 100% identity to the IS10 element found in enterobacteria such as *E. coli* and *Shigella flexneri*. To our surprise (and chagrin) many other sequences in the database from diverse sources such as *Homo sapiens*, *Drosophila*, *Arabidopsis*, and *Picea abies* chloroplast DNA contain IS10 (Table 1). Alignment of this sequence with the two inverted repeats of the complex Tn10 transposable element (Halling et al. 1982) indicated that this sequence was identical to the 3' end repeat of Tn10 and, therefore,

was identified as *IS10R*. The alignment also identified the expected 9-bp direct duplication (Halling and Kleckner 1982) flanking the insertion. The observed duplication contained a deviation from the reported consensus sequence for insertion:

Consensus 5'nGCTnAGCn3'
 BnBAC55::IS10 5'aGCTaTGCt3'.

Deviations from the consensus have been observed at each position by others (Kleckner et al. 1996). Inspection of the sequences listed in Table 1 also revealed the presence of the 9-bp direct repeat with variants at every position.

Figure 1C shows that only the mutant BnBAC55::*IS10R* has a 4.1-kb fragment that hybridizes with the *IS10R* probe. A weakly hybridizing fragment can be detected in the BnBAC25 lane in Fig. 1C. This signal could be due to contaminating *E. coli* chromosomal DNA in the BAC DNA preparation, or to an incipient mutant population carrying a newly inserted copy of *IS10R*.

Figure 1D presents a model to explain the results observed in panels A, B and C. BnBAC25 and BnBAC55 are overlapping clones of a contig. A 2.8-kb *EcoRI* fragment that contains BAC25L is present in the overlapping clones. The insertion of *IS10R*, which lacks an *EcoRI* site, is postulated to have occurred approximately 300 bp from the first *EcoRI* site downstream from the BAC25L sequence. This insertion causes the formation of a 4.1-kb *EcoRI* fragment that hybridizes with both BAC25L and the truncated *IS10R* probe.

Source of *IS10R*

To investigate the source of *IS10R*, we analyzed the *E. coli* strain DH10B, host for the *P. vulgaris* BAC library. Cells of this strain were grown from the same freezer stock (electrocompetent cells obtained from the manufacturer) used for the construction of the library, and genomic DNA was isolated for PCR and Southern-hybridization analysis. PCR-amplification was carried out with nested primers derived from the sequence obtained from BnBAC55::*IS10R*. A clear PCR amplification product of the predicted truncated size was obtained in these assays. Furthermore, Southern-blot hybridization of DH10B genomic DNA with the truncated *IS10R* amplification product revealed several hybridizing fragments. These results indicated that DH10B was the source of *IS10R*. This observation prompted us to investigate whether DH10B was harboring the typical *Tn10* transposon that carries the tetracycline resistance gene. To test this hypothesis, cells of BnBAC25, BnBAC55 and BnBAC55::*IS10R* were grown in LB containing 25 µg/ml of chloramphenicol, 6 or 12.5 µg/ml of tetracycline, and combinations of each tetracycline concentration with chloramphenicol. Growth was observed only when tetracycline was absent, suggesting that *IS10R* alone, and not the typical *Tn10*, was present in the cells.

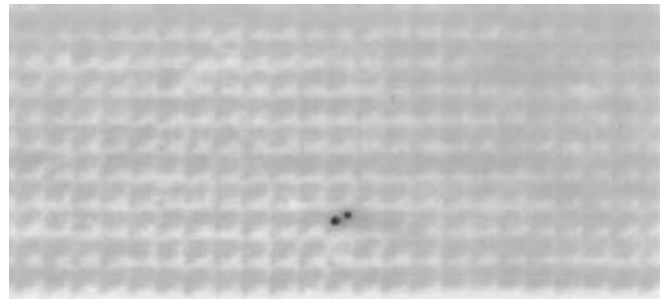


Fig. 2 Sample BAC filter-hybridization results with the *IS10R* sequence as a probe. Intensely hybridizing clones are observed with a frequency of 10^{-4} . BAC DNA prepared from the intensely hybridizing clones possesses a common restriction fragment that contains part of the *IS10R* sequence

Survey of the *P. vulgaris* BAC library

The entire BAC library (33000 clones) was screened with the truncated *IS10R* amplification product. Two levels of hybridization signal intensity were observed (Fig. 2). Clones exhibiting the strongest hybridization signals appeared to contain *IS10R* in the BAC DNA, whereas those exhibiting the weaker hybridization signal contained only the genomic copies. BAC DNA isolated from the intensely hybridizing clones, arising in approximately 1/10000 clones, contained a 2.3-kb *EcoRV* fragment that strongly hybridized to the *IS10R* probe (data not shown).

To better-characterize transposition in clones of the BAC library, we selected at random 30 BAC clones for analysis after an overnight culture and a subset of these after 8 days of daily subculturing. *EcoRV* digestions and Southern-blot hybridization with the truncated *IS10R* probe clearly showed the presence of five distinct fragments in each one of the 30 samples. These fragments were deemed to contain *IS10R* copies of genomic origin because they were identical to those obtained with genomic DNA from untransformed DH10B cells. These results are consistent with the observation that BAC DNA isolated with the miniprep procedure contained small amounts of chromosomal DNA contamination. Additional bands were also observed and probably represented recent transposition events. Figure 3 A shows the results for 15 of the 30 BAC DNA preparations. Ten of these clones were immediately subcultured on a daily basis for a period of 8 days. At the end of this period BAC DNAs were analyzed to determine whether *IS10R* had transposed. Eight of these samples are shown in Fig. 3B. Southern-blot hybridizations with the subcultured clones revealed the disappearance of some fragments (compare clones 4 and 7 from Fig. 3A and B), and the appearance of several new bands of varying intensity. Among these, a 2.3-kb *EcoRV* hybridizing fragment was noticed in each clone. This new fragment co-migrated with that contained in the BAC DNA isolated from the clone (30G19) exhibiting strong hybridization to *IS10R* in Fig. 2.

The presence of the 2.3-kb *EcoRV* fragment in all clones after prolonged subculturing suggested that *IS10R*

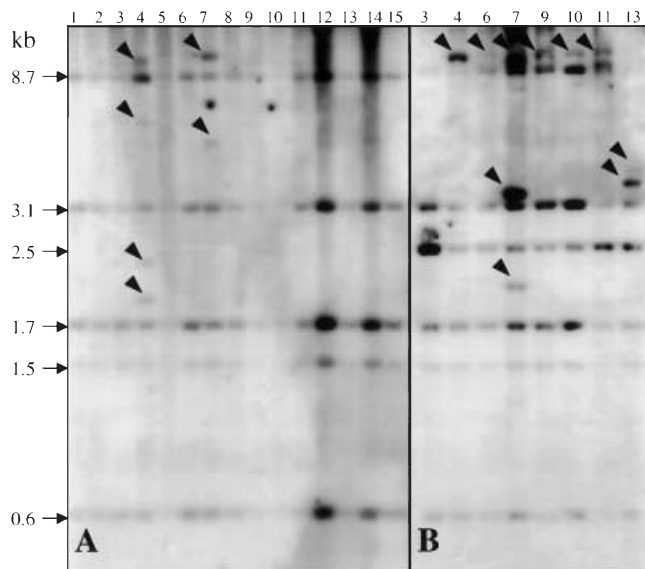


Fig. 3 Detection of *IS10R* transposition within selected *P. vulgaris* BAC clones. *Panel A* displays results of Southern-blot hybridization of the *IS10R* probe to BAC DNA from 15 randomly selected BAC clones. Shared fragments of 8.7, 3.1, 1.7, 1.5 and 0.6 kb represent non-uniform levels of contaminating host-genomic DNA. *Arrow heads* designate polymorphic fragments presumed to represent genomic rearrangements associated with *IS10R* activity. *Panel B* displays the results of a Southern-blot hybridization to BAC DNA from eight selected BAC clones following a series of eight daily sub-cultures. *Lane numbers* correspond to those in panel A. Note the uniform acquisition of a 2.3-kb fragment in all clones following sub culture

has a preferred site for transposition in the cloning vector. To test this hypothesis, we took two parallel approaches to determine the sequences flanking the transposon from BAC clone 30G19 shown in Fig. 2. First, the 2.3-kb *EcoRV* fragment was cloned into the pGEM-5Zf (Promega Corp) plasmid and the ends of the insert were sequenced. Second, *IS10R*-derived oligonucleotides were used as sequencing primers for the flanking regions. Our analysis revealed that the preferred sequence for transposition corresponds to position 1470 to 1478 of pBeloBAC11 (gi|1817728). This sequence also shows deviations from the consensus sequence:

Consensus 5'nGCTnAGCn3'
BAC 5'cGCCcGGTa3'.

Although three of the six bases deviate from the consensus, the internal four bases are in an inverted-repeat configuration. Transposition at this site does not appear to disrupt any of the vector functions because it is located in an intergenic region upstream from the CM^r gene. The 2.3-kb *EcoRV* fragment includes 265 bp from the 3' end of *IS10R* and 2027 bp from the BAC vector.

Discussion

A prokaryotic element, *IS10R*, was detected in a *P. vulgaris* BAC library and, by a homology search, in several

database sequence entries from diverse eukaryotic sources that include humans, *Arabidopsis*, yeast and *Drosophila*, among others. The natural presence of a highly conserved prokaryotic transposable element in diverse sequences of eukaryotic organisms seems highly unlikely and is believed to be a contaminant originating from the *E. coli* strain DH10B. The evidence to support this assertion is based on the observation that the majority of the *IS10R*-contaminated eukaryotic sequences detected in the databases were from BAC clones, and the fact that the *E. coli* DH10B strain is commonly used as a host for BAC libraries. However, more-definitive evidence was provided by hybridization and PCR-based detection of this element in the DH10B stock (obtained from Life Technologies, Inc) used as a host for the *P. vulgaris* BAC library. The detection of 9-bp direct repeats flanking *IS10R* in the database entries removes the possibility that sequence contamination could have occurred *in silico*. In some instances, the contaminated sequences have been annotated with regards to the presence of *IS10*, but the contaminating sequence has not been filtered out. This indicates that, although some researchers are aware of the contamination, the extent and origin of *IS10R* contamination is not widely known in the scientific community. To our knowledge, this contamination has not been reported in the literature.

Tn10 is a complex prokaryotic transposon comprising two inverted repeats of *IS10* flanking a tetracycline resistance gene (Halling et al. 1982). Sequence analysis indicated that the IS element detected in BnBAC55 corresponds to *IS10R*, the copy present at the 3' end of Tn10. This copy is an autonomous mobile element because, in contrast to *IS10L*, it has a functional transposase (Halling et al. 1982). The analysis presented here suggests that *IS10R* alone is present in the *E. coli* DH10B strain used in the construction of the *P. vulgaris* BAC library. This is probably true for the other BAC libraries. The presence of multiple hybridizing fragments on a Southern blot of genomic DNA of untransformed DH10B can be ascribed to the presence of multiple copies of the element in the DH10B genome. Eichenbaum and Livneh (1995) observed that transposition of *IS10R* was usually accompanied by an increase in the number of fragments that hybridized with the *IS10R* probe. The proposed mechanism of transposition is based on the fact that *IS10R* transposition is controlled by *dam* methylation. It is hypothesized that excision occurs in one of the hemimethylated sister molecules soon after the replication fork passes the *IS10* element. The donor molecule is degraded, and an insertion may occur at a point where the replication fork has not yet passed. As a result, the surviving sister molecule will carry two copies of the element (Bender et al. 1991).

The experiments presented here showed that *IS10R* can transpose into BAC DNA (vector or insert) with continuous subculturing. The existence of a preferred site for *IS10R*-transposition in the BAC vector is strongly suggested by the presence of an *IS10R*-hybridizing 2.3-kb *EcoRV* fragment in 30G19, and the appearance of

a similar fragment in all BAC clones after daily subculture. Transposition of *IS10R* upstream from CM^R in 30G19 appears to have occurred soon after transformation because the population is uniform, whereas the 2.3-kb fragment detected after daily subculturing is not as strong as other hybridizing fragments (Fig. 3B) and is assumed to result from a more-recent transposition event. The strong hybridization signal produced by 30G19 can be explained by a polarity effect (Szeverenyi et al. 1996). It has been reported that transposition of IS elements upstream from ORFs can bring about increased levels of expression. An increased level of expression of CM^R could give a selective advantage and result in increased growth and a higher hybridization signal. At the present time, we do not know whether transpositions into the cloned DNA are intermediary stages before the element ends up at the preferred vector site upstream from the chloramphenicol resistance gene.

The presence of *IS10R* within the genome of the host commonly used for BAC libraries presents serious problems. Perhaps the most obvious of them is the detection of false homologies after transposition into the insert, as indicated earlier. However, this problem can be overcome by appropriately filtering BAC sequences before submission. Another problem can be encountered during fingerprinting of BACs when the mutant population(s) increases to a significant level. More troublesome is the potential problem of alterations in the cloned sequences that can vary from short duplications to deletions and inversions (Kleckner et al. 1991; Binns 1993) and co-integrate formation in $RecA^+$ strains (Eichenbaum and Livneh 1995). We have only monitored the fate of a small sample of clones after continuous daily subculturing for a short period of time. A more detailed analysis of BAC clones is needed before the long-range fate of *IS10R*-contaminated BAC libraries can be ascertained. BAC clones from libraries known to be contaminated with *IS10R* can still be utilized as long as some properties of this element are taken into account. The rate of transposition of *IS10* has been estimated at 10^{-4} per element per generation (Kleckner et al. 1996), and transposition of this element is negatively regulated by *dam* methylation (Roberts et al. 1985). However, continuous incubation at the stationary phase (Skaliter et al. 1992), as well as exposure to UV radiation (Eichenbaum and Livneh 1998), can increase the rate of *IS10* transposition. Taking into account the properties of the transposable element during the handling of BAC libraries or individual clones may avoid or ameliorate the negative effects of the element. It is obvious that a new host free of transposable elements is needed before the construction of new BAC libraries is considered.

Acknowledgments We thank Dr. L.C. Hannah and Dr. K. C. Cline for reviewing this manuscript and offering some valuable suggestions. This work was supported in part by USDA grant No. 97121514 to C. E. V. and NIH grant R21 GM54154 and NSF grant IBN 9728329 to S. A. M. Florida Agricultural Experiment Station Publication No. R-07583. All the experiments described here were performed in full compliance with USA laws.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Bender J, Kuo J, Kleckner N (1991) Genetic evidence against intramolecular rejoining of the donor DNA molecule following *IS10* transposition. *Genetics* 128:687–694
- Binns M (1993) Contamination of DNA database sequence entries with *Escherichia coli* insertion sequences. *Nucleic Acids Res* 21:779
- Binns MM, Bournsnel MEG, Tomley FM, Brown TDK (1986) Comparison of the spike precursor sequences of coronavirus IBV strains M41 and 6/82 with that of IBV Beaudette. *J Gen Virol* 67:2825–2831
- Chomczynsky P (1992) One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Anal Biochem* 201:134–139
- Eichenbaum Z, Livneh Z (1995) Intermolecular transposition of *IS10* causes coupled homologous recombination at the transposition site. *Genetics* 140:861–874
- Eichenbaum Z, Livneh Z (1998) UV light induces *IS10* transposition in *Escherichia coli*. *Genetics* 149:1173–1181
- Gonzalez IL, Sylvester JE (1997) Incognito rRNA and rDNA in databases and libraries. *Genome Res* 7:65–70
- Halling SM, Kleckner N (1982) A symmetrical six-base pair target site sequence determines *Tn10* insertion specificity. *Cell* 28:155–163
- Halling SM, Simons RW, Way JC, Walsh RB, Kleckner N (1982) DNA sequence organization of *IS10*-right of *Tn10* and comparison with *IS10*-left. *Proc Natl Acad Sci USA* 79:2608–2612
- Jha PK, Sarkar S (1997) DNA sequencing and comparative sequence analysis reveal that the *Escherichia coli* genomic DNA may replace the target DNA during molecular cloning: evidence for the erroneous assembly of *E. coli* DNA into database sequences. *Comp Biochem Physiol* 118:333–339
- Kleckner N, Bender J, Gottesman S (1991) Uses of transposons with emphasis on *Tn10*. *Methods Enzymol* 204:139–180
- Kleckner N, Chalmers RM, Kwon D, Sakai J, Bolland S (1996) *Tn10* and *IS10* transposition and chromosome rearrangements: mechanism and regulation in vivo and in vitro. In Saedler H, Gierl A (eds) *Transposable elements*. Springer, Berlin, pp 49–82
- Miller C, Gurd J, Brass A (1999) A rapid algorithm for sequence database comparisons: application to the identification of vector contamination in the EMBL databases. *Bioinformatics* 15:111–121
- Plyler TR, Vallejos CE (2000) A method for cloning restriction fragments containing the termini of BAC inserts. *BioTechniques* 28:1012–1018
- Roberts D, Ascherman D, Kleckner N (1991) *IS10* promotes formation of adjacent deletions at low frequency *Genetics* 128:37–43
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Seluja GA, Farmer A, McLeod M, Harger C, Schad PA (1999) Establishing a method of vector contamination identification in database sequences. *Bioinformatics* 15:106–110
- Skaliter R, Eichenbaum Z, Shwartz H, Ascarelli-Goel R, Livneh Z (1992) Spontaneous transposition in the bacteriophage λ *cro* gene residing on a plasmid. *Mutat Res* 267:139–151
- Southern TM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517
- Szeverenyi I, Hodel A, Arber W, Olsas F (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* 174:103–110
- Vanhouten W, Mackenzie SA (1999) Construction and characterization of a common bean bacterial artificial chromosome library. *Plant Mol Biol* 40:977–983