Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization

David J.Earp[1], Brenda Lowe[1] and Barbara Baker[1,2]*
[1]Plant Gene Expression Center, University of California, Berkeley/USDA 800, Buchanan St., Albany, CA 94710 and [2]Department of Plant Pathology, University of California, Berkeley, CA 94720, USA

Received February 21, 1990; Revised and Accepted April 30, 1990

ABSTRACT

The isolation of sequences flanking integrated transposable elements is an important step in gene tagging strategies. We have demonstrated that sequences flanking transposons integrated into complex genomes can be simply and rapidly obtained using the polymerase chain reaction. Amplification of such sequences was established in a model system, a transgenic tobacco plant carrying a single *Ac* element, and successfully applied to the cloning of a specific *Spm* element from a maize line carrying multiple *Spm* hybridizing sequences. The described utilization of methylation sensitive restriction enzymes (including those with degenerate recognition sequences) in the generation of templates for amplification will simplify the cloning and mapping of genomic sequences adjacent to transposable elements.

INTRODUCTION

The utility of transposable elements as tools for the molecular tagging of genes, whereby insertion of a transposon causes a recognizable phenotypic change, has been elegantly demonstrated in prokaryotes and in several eukaryotic systems including *Drosophila* (1) and the mouse (2). In plants, where transposable elements were first discovered (3), native transposons have been used in the isolation of genes from maize (reviewed in 4) and *Antirrhinum* (reviewed in 5). However, in most plant species, transposons suitable for such tagging studies have yet to be identified and thus the utilization of well characterized elements in heterologous systems is being investigated.

Transposition of the maize controlling element *Activator* (*Ac*) (6) in several dicotyledonous plant species has recently been reported (7, 8, 9, 10, 11, 12). Subsequent characterization of this element in tobacco (8, 13 and Hehl and Baker, submitted) supports the suggested application of an *Ac* based system for gene tagging in transgenic plants (7). Sequences flanking transposed elements will additionally provide a new source of restriction fragment length polymorphism (RFLP) probes which will aid in the mapping of plant genomes. Towards these goals, chromosome mapping and tagging strategies for use in several plant species are being developed using *Ac* based tagging vectors.

The cloning of transposed elements and flanking sequences from transformed plants is an important step in these mapping and tagging studies. Conventionally the initial step in this cloning process has been achieved by constructing bacteriophage lambda libraries and screening these with element derived probes. To obviate the necessity of making and screening a library for each plant to be analyzed, we have utilized a polymerase chain reaction (PCR) based methodology (14) to amplify and clone sequences flanking transposed elements. The original PCR procedure permits amplification of regions of DNA between stretches of known sequence. A modification of this approach, termed inverse PCR (IPCR), allowing amplification of regions flanking segments of known sequence, has been described (15). We report here a detailed assessment of the feasibility of cloning genomic sequences flanking transposable elements from plant genomes by IPCR, factors affecting the efficacy of this approach, and experiments verifying the authenticity of amplified sequences.

A line of transgenic tobacco containing a single *Ac* element, *Ac-18*, was utilized as an experimental system. *Ac-18* is capable of inducing the transposition of *Ds* elements but is itself stable due to the deletion of 4 bp from one terminal inverted repeat sequence. The element transposed from an integrated *Agrobacterium* T-DNA vector sequence after transformation and inserted at a chromosomal location designated *Nt-1*. *Nt-1* is part of an undefined transcriptional unit and is comprised primarily of unique sequence (13; Hehl and Baker, submitted). Two approaches taken in order to obtain genomic sequences adjacent to *Ac-18* by IPCR are presented and the utility of methylation sensitive enzymes for the generation of IPCR templates is demonstrated. Specific amplification of sequences flanking the *Ac-18* element was achieved and up to 1 μg of specific product was readily obtained in one round of amplification (35 cycles) from 2–4 μg of genomic DNA.

Several genes have already been tagged and cloned from maize using the transposon families *Activator–Dissociation* (*Ac-Ds*) (16, 17, 18, 19, 20), *Suppressor Mutator* (*Spm*) or its equivalent, *Enhancer* (*En*), (21, 22, 23, 24, 25) and Mutator (Mu) (21, 26,

* To whom correspondence should be addressed

27). In such instances, cloning of the tagged gene is complicated by the presence in the genome of complex arrays of defective and cryptic elements hybridizing to element derived probes (21, 28). Digestion with methylation sensitive restriction enzymes to preferentially release active transposons from genomic DNA has been used in some cases to facilitate the cloning of inserted elements (17, 19, 22, 25). To ascertain the feasibility of using IPCR to clone specific elements and flanking sequences from such complex backgrounds, we applied the approach to amplify a sequence flanking a specific *Spm* element from a line of maize obtained from a screen for transposon induced mutants (Lowe and Hake, unpublished). An individual plant recovered from the screen was found to contain at least one newly transposed *Spm* element absent from progenitor and sibling plants; genomic sequences flanking this element were obtained by IPCR.

MATERIALS AND METHODS

General molecular biology techniques

EcoRII was obtained from B.R.L. and used in the supplied buffer. All other restriction enzymes and T4 DNA ligase were obtained from New England Biolabs. Digests were performed in the salt buffers described by Maniatis *et al.* (29) with the addition of bovine serum albumin (Boehringer, nucleic acid grade) to a final concentration of 100 $\mu g\ \mu l^{-1}$. Ligations were performed at 16°C for 12−24 hours in Biolabs' recommended buffer.

DNA fragments were purified for cloning purposes or for use as probes by electroelution from agarose gels onto NA45 paper (Schleicher and Schull) followed by butanol extraction, phenol/chloroform extraction and ethanol precipitation.

Template preparation

High molecular weight genomic DNA was extracted from *Nicotiana tabacum* by a modification of the published procedure (30) (S. Tanksley, personal communication, described in Hehl and Baker, submitted). The DNA obtained was further purified by centrifugation through cesium chloride gradients and ethanol precipitation (31). A simplified restriction map of the tobacco locus *Nt-1::Ac-18* is shown in Figure 1 together with an illustration of the principle by which one type of IPCR template was generated. In this instance the scheme permits amplification of sequences flanking both ends of the element. Following PstI digestion, tobacco DNA was extracted consecutively with phenol: chloroform (1:1) and chloroform: isoamyl alcohol (24 : 1) and then ethanol precipitated. After resuspension, this material was self ligated at a DNA concentration of approximately 2 $\mu g\ ml^{-1}$ and 1 unit μl^{-1} T4 DNA ligase. Subsequently the ligated material was extracted as above and precipitated with 3 volumes of ethanol in the presence of 0.1 volumes of 3M sodium acetate pH 5.6 and 2 $\mu g\ ml^{-1}$ tRNA. This material was then linearized by digestion with EcoRI followed by a further round of extractions and ethanol precipitation.

Genomic DNA was isolated from maize by the procedure of Shure *et al.* (32) digested with SalI and size fractionated on a continuous glycerol gradient. Fractions enriched for the 10.2 kb fragment were precipitated with isopropanol, ligated as described above and subsequently linearized with KpnI.

Oligonucleotides

Oligonucleotide primers were synthesized on an Applied Biosystems model 380B DNA synthesizer. Primers utilized in this work for amplification of sequences flanking *Ac* were:

A3 :- 5′ ATACGATAACGGTCGGTACGGG 3′
A4 :- 5′ TCGGGTTCGAAATCGATCGGGAT 3′
B5 :- 5′ CGTTTTCGTTTCCGTCCCGCAAG 3′
C1 :- 5′ GGCCAGCTCCAACTCCCATCCA 3′
C2 :- 5′ CATAACCACTGAGGCCCAGCCC 3′

Other primer sequences used to amplify flanking regions are available upon request.

Primer sequences for amplification of regions flanking *Spm* were:

L1 :- 5′ TCTCTCCAGTCTAGCCAGCGGCGT 3′
R3 :- 5′ GTGTGAGAATGACTTTATTCGGCT 3′

Polymerase chain reaction amplification

PCR reactions were performed using 2.5 units of Taq polymerase (Perkin Elmer Cetus), 0.2μM of each primer and 200μM of each dNTP in the 1.5mM $MgCl_2$ buffer recommended by the manufacturers. Amplification was achieved with a cycle of 1 min. denaturation at 94°C, 2 min. annealing at 55°C and 1−3 min. polymerase extension at 72°C repeated 35 times, with a final 6 min. 72°C extension step and a slow cool to 4°C using the Perkin Elmer Cetus Thermal Cycler. The products of the reactions were directly analyzed on 1% LE agarose (FMC) gels or 2.4% agarose gels comprising 2% Nusieve (FMC) agarose and 0.4% LE agarose.

Bacterial cloning hosts and vectors

The cloning vector M13KRV8.2 was derived from M13K8.2 (33) by mutagenesis and differed from M13K8.2 by the substitution of EcoRV sites for the SmaI sites (P. Carter, personal communication). This vector allows for efficient EcoK selection cloning of blunt ended fragments. *E. coli* JM101 ($r_k^+m_k^+$) (34) was employed for selection of clones bearing inserts.

Southern blot analyses

Genomic DNA was isolated and banded in cesium chloride gradients as described above. Following restriction enzyme digestion, 10 μg samples of DNA were electrophoresed in 0.8% agarose gels and subsequently transferred to nitrocellulose (Schleicher and Schull) (35). DNA fragments were radiolabelled by the random primer method using the Multiprime DNA Labelling Kit produced by Amersham. Labelled probes were hybridized to nitrocellulose filters according to Hughes *et al.* (36) at 42°C for 12−24 hours followed by three 15 min. washes in 0.1× SSC, 0.1% (w/v) SDS at 50°C.

RESULTS

Generation of IPCR template from tobacco genomic DNA

A scheme illustrating the production of template for IPCR amplification of sequences flanking *Ac* at the *Nt-1* locus in transgenic tobacco is illustrated in Figure 1. This template generation requires [1] PstI digestion of the genomic DNA followed by [2] intramolecular religation of the generated fragments and subsequently [3] linearization of ligated circles with EcoRI. Failure to achieve complete digestion at step [1] would result in several locus specific templates of various sizes while intermolecular rather than intramolecular religation at step [2] would produce various chimeric templates.

In order to assess the effects of the procedure on the genomic DNA, 10 μg aliquots of genomic DNA were removed after each step and analyzed by Southern blot hybridization using a labelled internal HindIII fragment of *Ac* as a probe (probe C, Figure 1).

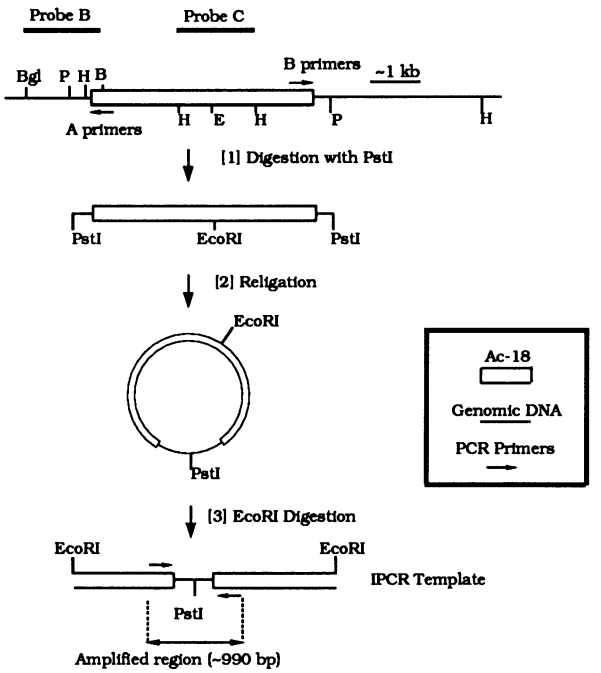

Figure 1. Schematic representation of IPCR template generation from genomic tobacco DNA carrying *Ac-18* inserted at the *Nt-1* locus. Probes B and C are further described in the text. IPCR primers not to scale. Restriction sites: Bgl–BglII; B–BamHI; P–PstI; E–EcoRI; H–HindIII.

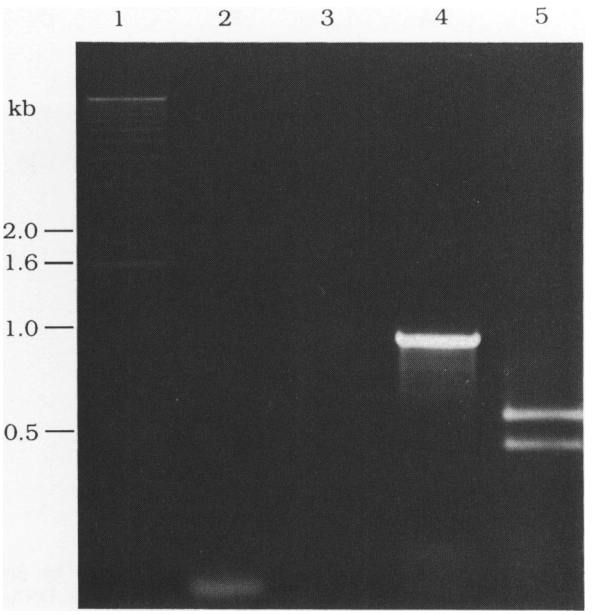

Figure 2. IPCR amplification of sequences flanking *Ac* at the *Nt-1* locus in tobacco with primers A4 and B5. Samples were electrophoresed in 0.8% agarose gel and visualized by ethidium bromide staining. Lane 1, 1kb ladder molecular weight markers (New England Biolabs); lane 2, amplification without template DNA (primers only); lane 3, amplification of undigested DNA from a plant homozygous for *Nt-1*:: *Ac-18*; lane 4, amplification of IPCR template generated from a plant homozygous for *Nt-1*:: *Ac-18* as shown in Figure 1; lane 5, amplified material as lane 4 digested with PstI.

The results confirmed the presence of an approximately 5.3kb fragment bearing the element in PstI digested genomic DNA (13) and detected a new product of slower mobility (at an equivalent of ca. 6.8kb) following ligation. After EcoRI digestion all hybridizing material migrated at 5.3 kb (not shown). This suggested that IPCR template should be generated from genomic DNA under the conditions employed.

IPCR amplification of genomic sequences flanking *Ac* in tobacco

Primers were prepared (see Methods) corresponding to each terminal region of the *Ac* sequence (37, 38, 39, 40) as 'nested primer' sets (41); primer pairs were selected such that one primer (from the group designated A primers in Figure 1) was homologous to the non-coding strand of *Ac*, allowing polymerase extension out of the element through the 5′ end, and the second primer (from the B primers, Figure 1) was homologous to the coding strand at the 3′ end of the template molecule.

Approximately 4μg of the tobacco IPCR template material were subjected to the polymerase chain reaction as described in Methods, with a 3 minute polymerase extension step. After 35 cycles 10% samples of the product were removed and subjected to electrophoresis either directly or after digestion with PstI; the presence of a PstI site within the product is diagnostic for amplification of the desired region (Figure 1). All primer pairs from groups A and B produced amplified fragments of the expected size (ca. 1 kb) containing a PstI site (variations in size that were observed with different primer pairs correlated with differing locations of primers). Amplification reactions with primers alone and on untreated tobacco DNA produced no visible products under the given conditions. Figure 2 shows an agarose gel of the products typical of such reactions; in this instance primers A4 and B5 were employed to generate a product of ca. 990 bp (lane 4). The amplified sequence was cleaved by PstI to yield two products of predicted sizes ca. 450 and 550 bp (lane 5). A yield of ca. 1μg of the amplified product was typically obtained from these reactions.

Confirmation of IPCR product identity by Southern blot analysis

Nicotiana tabacum is an amphidiploid species arising from the hybridization of the two diploid species *N. sylvestris* and *N. tomentosiformis* (42). *Ac* integrated into the *Nt-1* locus of the *N. tomentosiformis* genome within *N. tabacum* to create *Nt-1::Ac-18* (13). In *N tabacum* SR1 the wild-type *N. tomentosiformis Nt-1* locus is detected as an approximately 8.5 kb band upon hybridization of an *Nt-1* locus-derived probe to BglII digested genomic DNA. In transgenic *N. tabacum* SR1 homozygous for *Nt-1::Ac-18*, this locus is detected as a band of approximately 13 kb (13). These hybridization patterns were exploited to confirm the identity of the amplified material.

The ca. 990 bp amplified band obtained using primers A4 and B5 was purified by electroelution, ligated into EcoRV cut M13KRV8.2 and transformed into *E. coli* JM101. The cloning of blunt ended amplified sequences is preferable, in this situation, to utilization of sites built onto PCR primers since it does not require digestion of the uncharacterized amplified sequences. The utilization of primers with built-in sites identical to the second cut enzyme in the template preparation (in this instance EcoRI) would be feasible, however the choice of this second cut enzyme is subject to change with different starting samples due to variations in restriction sites of flanking regions. Cloned IPCR product was released from cesium chloride purified 'phage DNA by EcoRI–BamHI digestion, purified by electroelution and

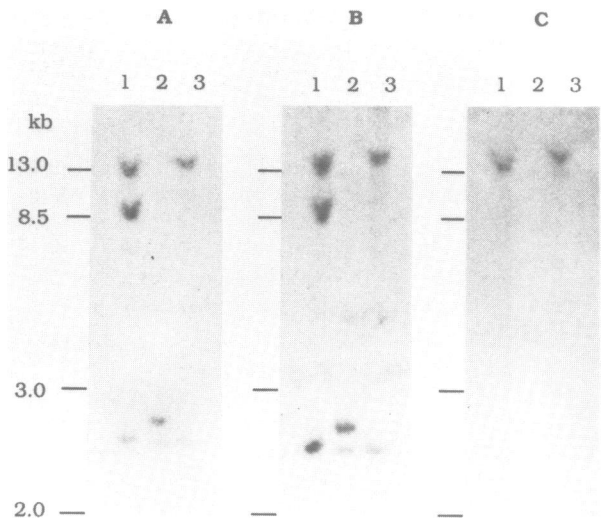

Figure 3. Confirmation of tobacco IPCR product identity by Southern hybridization. Southern blots of BglII digested tobacco genomic DNAs were hybridized with probes derived from: panel A, the cloned *Nt-1* locus ; panel B, the cloned IPCR product and panel C, the *Ac* element (probes are shown in Figure 1 and described in the text). DNA samples were extracted from: lane 1, transgenic SR1 tobacco heterozygous for the *Nt-1*:: *Ac-18* allele; lane 2, untransformed Samsun NN tobacco; lane 3, transgenic SR1 tobacco homozygous for the *Nt-1*:: *Ac-18* allele.

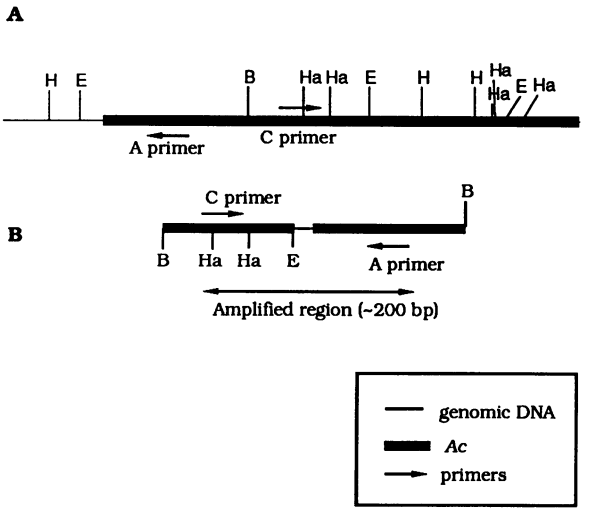

Figure 4. Schematic representation of IPCR template for amplification of genomic sequences adjacent to the 5' end of *Ac-18* at the *Nt-1* locus. **(A)** Partial restriction map of the 5' end of *Nt-1*:: *Ac-18* showing location of IPCR primers (primers not to scale) **(B)** IPCR template generated from the locus following digestion with EcoRII, religation and linearization with BamHI. Restriction sites: B−BamHI; E−EcoRII; H−HinfI; Ha−HaeIII

radiolabelled. Radiolabelled probes were also made from the internal 1.6 kb HindIII fragment of *Ac* and the 1.2 kb BglII−BamHI fragment of *Nt-1*::*Ac-18* , subcloned from a lambda library, (13) (probes C and B respectively, indicated on Figure 1).

Hybridization of these probes to BglII digests of tobacco genomic DNA is shown in Figure 3. *N. tabacum* SR1 homozygous for the *Nt-1*::*Ac-18* locus exhibits a 13 kb BglII fragment detected with both *Nt-1* and *Ac* derived probes (panels A and C, lane 3), and plants heterozygous for this allele (i.e. of genotype *Nt-1*::*Ac-18* / *Nt-1*) show this 13 kb band when hybridized to the *Ac* probe (panel C, lane 1) and an additional band of 8.5 kb representing the *Nt-1* allele when hybridized to the *Nt-1* probe (panel A, lane 1). The IPCR derived probe produced hybridization patterns, as shown in panel B, identical to the *Nt-1* probe, confirming IPCR amplification of the target locus flanking the transposed *Ac-18* element. Genomic DNA extracted from untransformed *N. tabacum* Samsun NN carries a BglII restriction fragment length polymorphism at the *Nt-1* locus (R. Hehl, unpublished) revealed by hybridization of both the *Nt-1* and IPCR probes to a fragment of ca. 2.7kb (panels A and B, lane 2). A band of ca. 2.5kb representing the *N. sylvestris Nt-1* homolog (13) was detected in both types of tobacco with the *Nt-1* homologous probes (panels A and B, lanes 1−3).

PCR amplification of genomic sequences flanking a single end of *Ac* in tobacco

Application of the methodology described above to the *Nt-1*::*Ac-18* locus produced amplification of sequences flanking both ends of the inserted transposon, and was aided by the prior knowledge of restriction sites within the flanking sequences. The number of enzymes suitable for the initial cut in this IPCR scheme is restricted to those which do not cleave within the *Ac* element and is further reduced to a sub-group of these with sites in the flanking genomic DNA relatively close to each end of the inserted element (within 5kb of each end, given the current limits of PCR amplification). In general application, suitable enzymes could be determined by the use of Southern blot analysis of the genomic DNA harboring transposons of interest. However, while the cloning of sequences flanking both ends of an element would be useful in studies of, for example, transposition target sequences, in situations where the goal is to obtain a probe for the target region, amplification of sequences flanking only one end is sufficient. This enables the utilization of the many frequent-cutting enzymes with sites within the element for the initial digestion. This notion was used to generate IPCR templates from a single end of *Ac* and adjacent genomic flanking sequences at the *Nt-1*::*Ac-18* locus. Figure 4A shows a restriction map of one end of *Ac-18* within the *Nt-1* locus and the location of primers used for this approach. Templates were generated using either HinfI or EcoRII for the first digestion, followed by ligation and then linearization with BamHI. Figure 4B shows the IPCR template molecule produced using EcoRII for the initial digestion step in this scheme. The template material obtained was subjected to PCR amplification using primers A4 and C1 as described in Methods with a 1 minute polymerase extension step.

Amplification of the material produced using HinfI for the initial digestion resulted in the synthesis of a continuous range of product sizes of ca. 305 bp (the predicted product size) and larger, visualized as a smear on an ethidium bromide stained agarose gel. Reamplification of this material with a nested primer pair (A3 and C2) produced a smear of slightly lower average molecular weight, again with a minimum size of the predicted product of amplification with this primer pair. Southern blot analysis using a probe derived from the *Nt-1* locus revealed that these smears hybridized to the probe (data not shown).

As discussed above, either incomplete initial digestion or intermolecular religation will produce a range of true IPCR templates of varying sizes containing priming sites and the desired target sequence but with additional, amplifiable sequences between the chosen initial restriction sites. Material from the

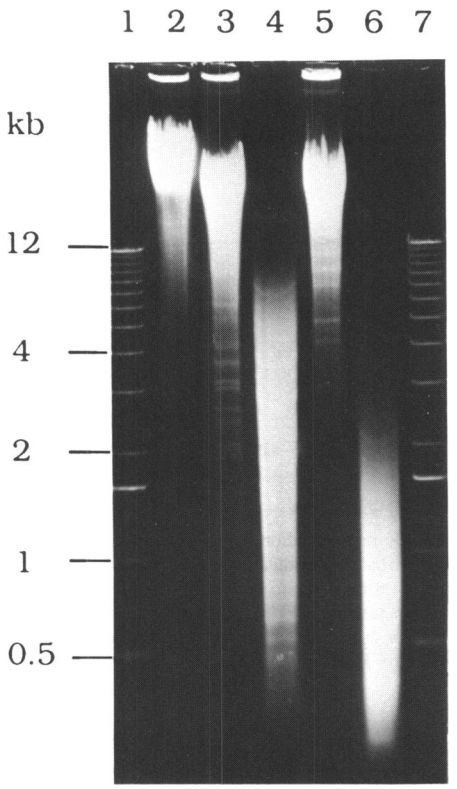

Figure 5. The effect of methylation on the digestion of tobacco genomic DNA. Tobacco DNA was digested, electrophoresed in 0.8% agarose and stained with ethidium bromide. Lanes 1 and 7, 1kb ladder molecular weight markers; lane 2, undigested genomic DNA; lanes 3–6 genomic DNA digested with EcoRII, BstNI, PstI and HinfI, respectively.

smears produced from the HinfI derived template was cloned into M13KRV8.2 and inserts of varying size were dideoxy sequenced (43). Amplified sequences arising following incomplete initial digestion would carry recognizable *Nt-1* locus sequences beyond the HinfI sites whereas templates created by intermolecular ligation would contain random genomic HinfI fragments. Sequence analyses showed no evidence of products resulting from partial digestion and sequences of three clones suggested that intermolecular ligation was indeed occurring.

The frequency of intermolecular ligation in the template preparation procedure will increase as the concentration of ligatable ends in the reaction increases. In turn the concentration of ends will depend on the frequency of cleavage in the preceding restriction digest. The products of various restriction enzyme digestions on tobacco DNA are shown in Figure 5. Digestion of tobacco DNA with HinfI produces numerous small length products (with an average size of 0.5–1kb, lane 6). Because of the concentration of ends produced by such frequent digestion, the probability of intermolecular religation following HinfI digestion is likely to be greater than that produced by ligation of, for example, the products of PstI digestion which cleaves less frequently (lane 5). This situation may be encountered with many frequent cutting enzymes; ligation of template material at lower DNA concentrations may alleviate the situation.

We chose to utilize previous observations of methylation of *Ac* elements and flanking sequences to approach the problem. In maize, active *Ac* elements and flanking sequences are undermethylated compared with inactive elements and total genomic DNA (44, 45). Similarly *Ac* and flanking sequences in

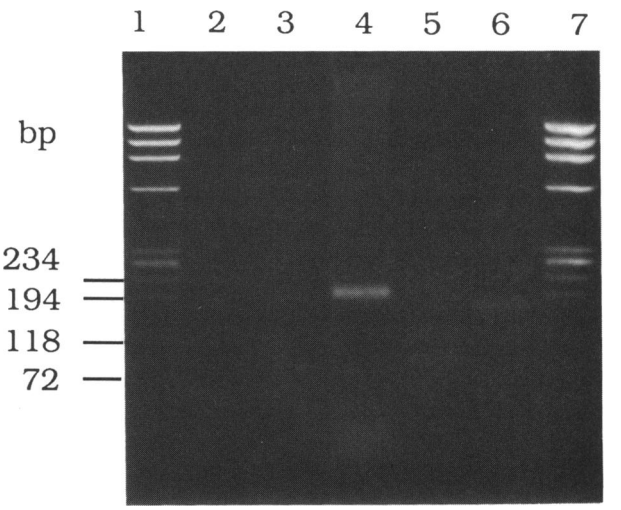

Figure 6. IPCR amplification of sequences flanking the 5′ end of *Ac-18* at the *Nt-1* locus in tobacco with primers A4 and C1. Samples were electrophoresed in 2.4% agarose gel and visualized by ethidium bromide staining. Lanes1 and 7, phi X174 HaeIII digest; lane 2, amplification without template DNA (primers only); lane 3, amplification of undigested DNA from a plant homozygous for *Nt-1*:: *Ac-18*; lane 4, amplification of IPCR template generated from a plant homozygous for *Nt-1*:: *Ac-18* as shown in Figure 4; lanes 5 and 6, amplified material as lane 4 digested with EcoRII and HaeIII respectively.

tobacco and tomato are apparently undermethylated relative to total genomic DNA (B. Baker, unpublished). Thus the use of methylation sensitive enzymes could potentially achieve the desired effect of frequent digestion within and around the element while leaving much of the genomic DNA in relatively large pieces. EcoRII sites within and around active *Ac* elements are only partially methylated in maize (45) and genomic EcoRII sites are generally methylated. The effect of methylation on EcoRII digestion of tobacco genomic DNA can be seen by comparison with the digest produced by the methylation insensitive isoschizomer BstNI (Figure 5, lanes 3 and 4 respectively). We subsequently used EcoRII to generate the IPCR template illustrated in Figure 4B.

Figure 6 shows the products of amplification obtained using primers A4 and C1 on IPCR template produced using EcoRII for initial digestion. A single band of ca. 200 bp was obtained from the IPCR template (lane 4) and no bands were observed using these primers without template or with untreated DNA (lanes 2 and 3). The 200 bp product contained the predicted EcoRII/BstNI site (lane 5) however, complete digestion of the product with these enzymes could not be achieved. The possibility of an underlying non-specific product was ruled out by the complete digestion of the material with HaeIII to yield fragments of ca.160 and ca. 35 bp (lane 6). Additionally, the ca. 200 bp product hybridized to the *Nt-1* locus probe (not shown).

The recognition sequences of both EcoRII (5′ CC A/T GG 3′) and HinfI (5′ GANTC 3′) are degenerate. The 5′ EcoRII site within *Ac* and the EcoRII site within the *NT-1* locus (Figure 4) differ at the central position of the recognition sequence and succesful amplification therefore requires ligation of cohesive ends carrying a one base pair mismatch. The ability of T4 ligase to perform such ligations has been previously described (46, 47, 48, 49). Similarly, sequence analyses of the HinfI sites contained in the amplified sequences cloned from the array of *Nt-1* locus IPCR products generated using HinfI for the initial digestion demonstrated that mismatch ligations and subsequent resolution of mismatches had occurred.

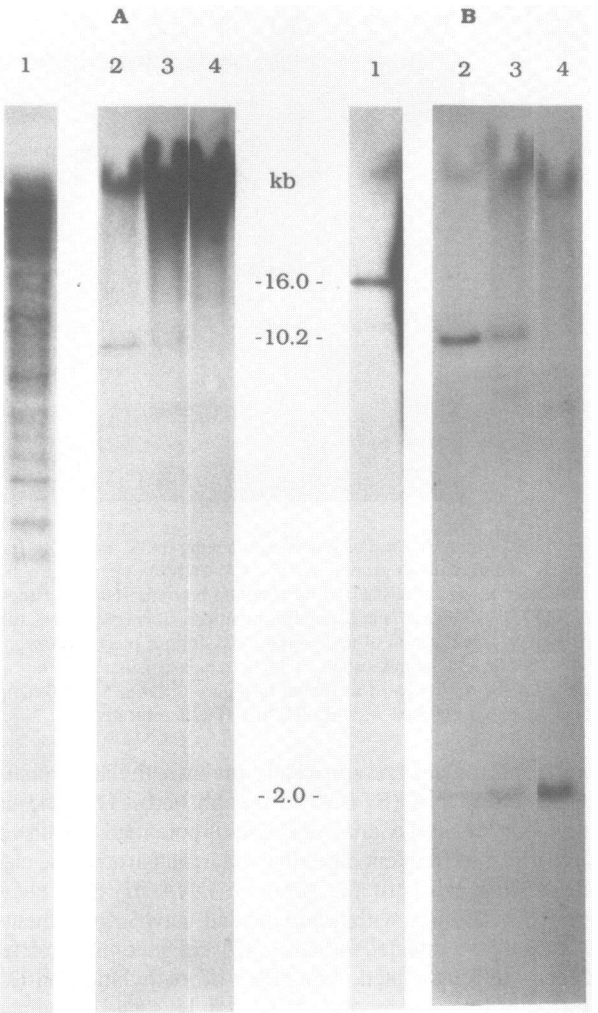

Figure 7. Southern blots of maize genomic DNAs. Southern blots of digested maize DNAs were hybridized with, in panel A, the *Spm* probe pBX1 and in panel B, the IPCR product from plant #554−27. Lane 1, DNA from the progenitor c2-m1 line digested with BamHI; lanes 2, 3 and 4 SalI digested DNA from F2 progeny resulting from outcrossing plant #554−27 and selfing the resultant F1 plants. The filter shown in panel A was stripped and rehybridized with the IPCR probe to give panel B.

IPCR amplification of sequences flanking *Spm* in maize

Following the successful amplification of sequences flanking *Ac-18* in this single element model system in tobacco, we wished to assess the utility of IPCR in situations where the cloning of a single, specific element from a complex background of homologous sequences is required. For this purpose we chose to amplify sequences flanking an *Spm* element in maize, which also allowed the testing of the approach in a second experimental system.

A maize line carrying the c2-m1 allele, a mutable allele of C2 arising from the insertion of an autonomous *Spm* element at the C2 locus, was originally described by McClintock (50). Figure 7, panel A, lane 1 shows a BamHI digest of genomic DNA from this maize line hybridized with the *Spm* probe pBx-1, the 450 bp BanII−XbaI fragment shown in Figure 8A. The numerous hybridizing bands illustrate the complexity of this line with regard to *Spm* homologous sequences.

A maize plant obtained by outcrossing the line carrying c2-m1

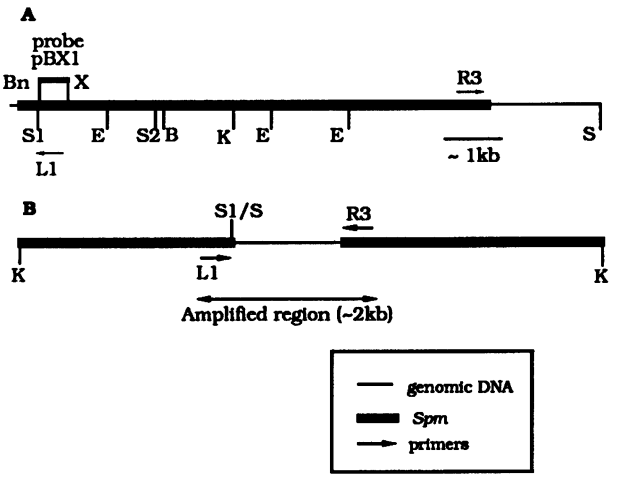

Figure 8. Schematic representation of IPCR template generated from genomic maize DNA. **(A)** Partial restriction map of *Spm* showing locations of pBX1 probe and IPCR primers (primers not to scale). The genomic SalI site 3′ of the *Spm* element selected as a target for IPCR is also indicated. **(B)** IPCR template derived from the selected *Spm* element shown in (A) indicating amplified region. Restriction sites: B−BamHI; Bn−BanII; E−EcoRI; K−KpnI; S, S1 and S2−SalI; X−XbaI.

was selected as a test system. This plant, #554−27, was found to contain at least one new *Spm* insertion, absent from the parent, which was characterized as a 10.2 kb SalI fragment hybridizing to the *Spm* probe pBX1. Figure 7, (panel A, lane 2) shows a progeny plant derived from #554−27 carrying the new 10.2 kb band. SalI is a methylation sensitive enzyme which cleaves maize genomic DNA only infrequently; the high molecular weight hybridizing material in Figure 7, panel A represents *Spm* homologous sequences not 'released' by the SalI digestion (see below). The *Spm* element identified by the 10.2 kb SalI fragment was selected as a target for IPCR amplification in order to obtain flanking genomic sequences.

The nucleotide sequence of *Spm* (51, 52) predicts an internal 2.1 kb SalI fragment that would be detected by the pBX1 probe (S1−S2, Figure 8A). However, Banks *et al.* (53) have shown that the SalI site at position 2373 (S2, Figure 8a) of *Spm* is always methylated and that methylation at the position 267 SalI site (S1, Figure 8A) is correlated with the activity of the *Spm* element; active elements being unmethylated, inactive elements being methylated at this site. It was assumed that the new 10.2 kb *Spm* hybridizing band was the result of insertion of an active *Spm* element and thus that the 10.2 kb fragment was generated by SalI cleavage at S1 in *Spm* and a flanking, genomic SalI site as shown in Figure 8A. Primers were therefore designed to prime synthesis out of the 5′ end of the element from the region immediately 3′ of the S1 SalI site (primer L1, Figure 8) and from the 3′ end of the element (primer R3).

IPCR template was generated by enriching SalI digested genomic DNA for the 10.2 kb fragment by size fractionation on a continuous glycerol gradient, followed by ligation and then linearization with KpnI. The resulting template is illustrated in Figure 8B. PCR amplification was performed as described for tobacco. Based on the restriction sites present in an autonomous *Spm* element, and the 10.2 kb fragment produced by SalI digestion of the genomic DNA, successful amplification would be expected to yield a ca. 2 kb product (Figure 8B). The result of PCR amplification of the template material with primers L1 and R3 is shown in Figure 9; five major products of

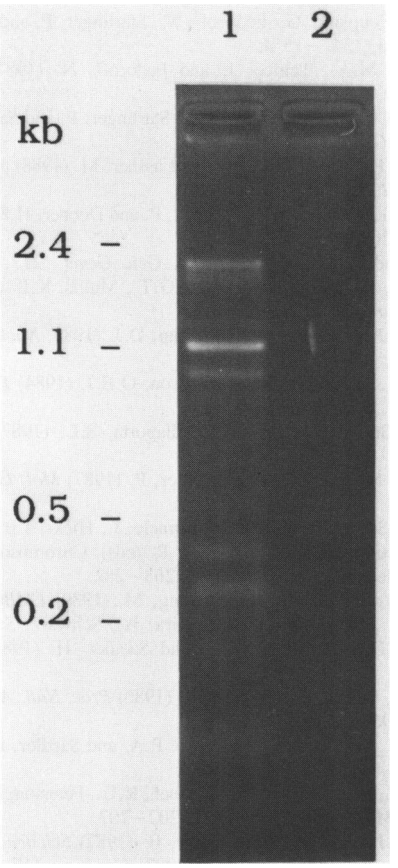

Figure 9. IPCR products obtained from maize DNA using primers L1 and R3. Samples were electrophoresed in 0.8% agarose gel and visualized by ethidium bromide staining. Lane 1, amplification with IPCR template generated from maize line #554−27; lane 2, amplification without template DNA (primers only).

approximately 2.0, 1.1, 1.0 , 0.4 and 0.3kb were obtained. The 2.0 kb IPCR product was purified, radiolabelled and hybridized to BamHI digested genomic DNA from the progenitor, c2-m1 carrying line. A single band, shown in Figure 7 (panel B, lane 1), of ca.16 kb was detected indicating that the amplified region represented unique sequence.

Subsequently, plant #554−27 was outcrossed and the resulting F1 plants were selfed. Genomic DNA was extracted from the F2 plants and digested with SalI. Figure 7, (panels A and B, lanes 2−4) shows hybridization of the 2.0 kb IPCR product and pBX1 to 3 representatives of this population, which segregated for the 10.2 kb SalI band. The pBX1 probe (panel A) detected a band of 10.2 kb in lanes 2 and 3; the IPCR probe (panel B) detected a band of 10.2 kb in lanes 2 and 3 and an additional band of 2.0 kb in lanes 3 and 4. A very weak 2.0 kb band was observed in lane 2, panel B. The IPCR probe also hybridized to the 10.2 kb and 2.0 kb SalI bands in plant #554−27 and the 2.0 kb band in SalI digested DNA from the progenitor line carrying c2-m1 (data not shown). The 2.0 kb fragment differs in size from the 10.2 kb band by approximately the size of the *Spm* element, which is ca. 8.3 kb. This confirms that the IPCR product did indeed represent regions flanking the *Spm* insertion originally detected as a 10.2 kb SalI fragment in plant #554−27. The hybridization data suggest that the 2.0 kb band (panel B, lanes 2−4) represents the target locus into which the *Spm* element inserted and so it is concluded that the DNA in lane 4, showing hybridization to only the 2.0 kb band, originated from a plant harboring no *Spm* insertion at this locus. The presence of a dark 10.2 kb band and a very light 2.0 kb band in panel B lane 2 suggests that this DNA originated from a plant homozygous for the *Spm* insertion, and that the faint 2.0 kb band represents somatic excision of the element from this locus. The presence of 10.2 and 2.0 kb bands of equal relative intensity in panel B lane 3 could be indicative of a plant heterozygous for the *Spm* insertion at this locus or of somatic excision of the element from the target locus.

A clone from a lambda library made from the #554−27 line was obtained using the IPCR product as a probe. Restriction mapping of this clone confirmed the presence of an *Spm* element with restriction sites identical to previously mapped autonomous *Spm* elements (51, 52) and the flanking genomic SalI site shown in Figure 8A. The IPCR product was also employed to determine the genomic location of this *Spm* insertion using the recombinant inbred lines described by Burr *et al.* (54). The amplified sequence (and therefore the transposed *Spm* element) mapped to position 112 on the long arm of chromosome 4, approximately 5 map units proximal to the C2 locus. The IPCR product which serves as an RFLP probe for this chromosomal location was designated pge1 (B. Burr, personal communication). The 1.1 and 1.0 kb products from the IPCR amplification were also mapped to single loci, unlinked to pge1. These and the smaller products, which were not mapped, may represent amplification of sequences flanking *Spm* related elements copurified with the 10.2 kb fragment on the glycerol gradient.

Sequences flanking a specific *Spm* element were thus readily amplified from a maize plant carrying numerous *Spm* homologous sequences.

DISCUSSION

We have demonstrated that sequences flanking transposons integrated into complex genomes can be simply and rapidly obtained using the polymerase chain reaction on suitably manipulated genomic DNA. Amplification of such sequences was established in a model transgenic tobacco plant carrying a single *Ac* element and successfully applied to the cloning of a specific *Spm* element from a maize line carrying multiple *Spm* hybridizing sequences. The method should provide considerable advantages in obtaining element flanking sequences compared with conventional lambda cloning and screening: for example, the screening of ca. one million clones of a maize library is typically necessary in order to yield a 95% chance of finding a single copy gene (55). IPCR amplification of genomic sequences flanking transposed elements will generate a new source of RFLP probes for use in genome mapping and in combination with the availability of cosmid and YAC libraries of tomato and *Arabidopsis*, the described methodology will greatly facilitate the isolation of wild type genes following tagging experiments.

The selection of restriction enzymes employed in the template generation procedure affected the ability to produce specific products. Previous studies have demonstrated that in maize, active *Ac* elements are undermethylated relative to the overall genome (44, 45). This work demonstrated the effectiveness of methylation sensitive restriction enzymes for IPCR template generation in the amplification of genomic sequences flanking unmethylated transposable elements in both tobacco and maize. The simultaneous amplification of genomic sequences adjacent to both ends of an integrated transposon requires, for template generation, the use of a restriction enzyme that does not cleave within the transposon and which cleaves the genomic DNA relatively close

to the element. Given these limitations, amplification of sequences flanking a single end of the element may be the approach of choice. Methylation sensitive enzymes may prove especially useful in this situation as they are able to cleave the unmethylated DNA in and around the element whilst cleaving the generally highly methylated genomic DNA only infrequently, alleviating problems of intermolecular ligation in subsequent stages. For the *Ac* element, the methylation sensitive enzyme EcoRII was found to be suitable for obtaining genomic sequences adjacent to the 5′ end of the element, while amplification of genomic sequences flanking one end of an *Spm* element from maize was achieved using SalI.

EcoRII and HinfI both have degenerate recognition sequences and at first sight would therefore appear to be poor choices for the creation of IPCR template due to the possible requirement for mismatch ligations. However we have clearly shown that such ligations were achieved with T4 ligase under normal conditions. The requirement for mismatch ligation with EcoRII did not result in a detectable decrease in yield of product after amplification when compared with the utilization of PstI or SalI for template generation. This observation extends the commonly considered repertoire of enzymes available for use in template generation. Up to 1 μg of specific product was readily obtained in one round of amplification (35 cycles) from 2−4μg of genomic DNA from the amphidiploid species *N. tabacum* (haploid genome size 2×10^6 kbp, (56)). This suggests that the method will be generally applicable to situations where large amounts of DNA cannot be obtained for instance from single, transformed *Arabidopsis* plants (haploid genome size 7×10^4 kbp, (57)).

The methodology has applications for the determination of chromosomal locations for integrated sequences other than transposons such as T-DNA integrated as a result of *Agrobacterium* mediated transformation and genes integrated following electroporation or particle bombardment. The wide potential usage of this approach in other systems is also apparent; the amplification of DNA flanking integrated ecotropic proviral sequences from BALB/c mice using similar methodology has recently been reported (58).

ACKNOWLEDGEMENTS

We are grateful to B. Osborne for critical advice and discussion, R. Hehl for materials and unpublished data on the *Nt-1* :: *Ac-18* allele, C. Waddell for critically reading the manuscript, S. Hake and J.A. Todd for discussion, B. Burr for analyzing maize recombinant inbred line data and J. Mathern for help with figure preparation. This work was supported by USDA Cris. # 5335-22230-002-00D to B. Baker and # 5335-22230-001-00D to S. Hake.

REFERENCES

1. Bingham, P.M., Levis, R. and Rubin, G.M. (1981) *Cell*, **25**, 693−704.
2. Rinchik, E.M., Russel, L.B., Copeland, N.G. and Jenkins, N.A. (1986) *Genetics*, **112**, 321−342.
3. McClintock, B. (1948) *Carnegie Inst. Wash. Year Book*, **47**, 155−169.
4. Fedoroff, N. (1989) In Berg, D.E. & Howe, M.M. (eds), Mobile DNA. American Society for Microbiology, Washington DC,375−411.
5. Coen, E.S., Robbins, T.P., Almeida, J., Hudson, A. and Carpenter, R. (1989) In Berg, D.E. & Howe, M.M. (eds), Mobile DNA. American Society for Microbiology, Washington D.C.,413−436.
6. McClintock, B. (1947) *Carnegie Inst. Wash. Yearbook*, **46**, 146−152.
7. Baker, B., Schell, J., Lorz, H. and Fedoroff, N. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 4844−4848.
8. Baker, B., Coupland, G., Fedoroff, N., Starlinger, P. and Schell, J. (1987) *EMBO J.*, **6**, 1547−1554.
9. Van Sluys, M.A., Tempe, J. and Fedoroff, N. (1987) *EMBO J.*, **6**, 3881−3889.
10. Knapp, S., Coupland, G., Uhrig, H., Starlinger, P. and Salamini, F. (1988) *Mol. Gen. Genet.*, **213**, 285−290.
11. Yoder, J.I., Palys, J., Alpert, K. and Lassner, M. (1988) *Mol. Gen. Genet.*, **213**, 291−296.
12. Jones, J.D.G., Carland, F.M., Maliga, P. and Dooner, H.K. (1989) *Science*, **244**, 204−207.
13. Hehl, R. and Baker, B. (1989) *Mol. Gen. Genet.*, **217**, 53−59.
14. Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) *Nature*, **324**, 163−166.
15. Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) *Nucl. Acids Res.*, **16**, 8186.
16. Fedoroff, N., Furtek, D.B. and Nelson, O.E.J. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 3825−3829.
17. Chen, J., Greenblatt, I.M. and Dellaporta, S.L. (1987) *Genetics*, **117**, 109−116.
18. Theres, N., Scheele, T. and Starlinger, P. (1987) *Mol. Gen. Genet.*, **209**, 193−197.
19. Dellaporta, S.L., Greenblatt, I., Kermicle, J., Hicks, J.B. and Wessler, S. (1988) In Gustafson, J.P. & Apples, R. (eds), Chromosome structure and function. Plenum Press, New York,263−282.
20. Hake, S., Vollbrecht, E. and Freeling, M. (1989) *EMBO J.*, **8**, 15−22.
21. O'Reilly, C., Shepherd, N.S., Pereira, A., Schwarz, S.Z., Bertram, I., Robertson, D.S., Peterson, P.A. and Saedler, H. (1985) *EMBO J.*, **4**, 877−882.
22. Cone, K.C., Burr, F.A. and Burr, B. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9631−9635.
23. Paz-Ares, J., Wienand, U., Peterson, P.A. and Saedler, H. (1986) *EMBO J.*, **5**, 829−834.
24. Wienand, U., Weydemann, U., Niesbach, K.U., Peterson, P.A. and Saedler, H. (1986) *Mol. Gen. Genet.*, **203**, 202−207.
25. Schmidt, R.J., Burr, F.A. and Burr, B. (1987) *Science*, **238**, 960−963.
26. McLaughlin, M. and Walbot, V. (1987) *Genetics*, **117**, 771−776.
27. Martienssen, R.A., Barkan, A., Freeling, M. and Taylor, W.C. (1989) *EMBO J.*, **8**, 1633−1640.
28. Fedoroff, N., Wessler, S. and Shure, M. (1983) *Cell*, **35**, 235−242.
29. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.
30. Bernatzky, R. and Tanksley, S.D. (1986) *Theor. Appl. Genet.*, **72**, 314−321.
31. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1988) Current protocols in molecular biology. New York, Greene and Wiley Interscience.
32. Shure, M., Wessler, S. and Fedoroff, N. (1983) *Cell*, **35**, 235−242.
33. Waye, M.M.Y., Verhoeyen, M.E., Jones, P.T. and Winter, G. (1985) *Nucl. Acids Res.*, **13**, 8561−8571.
34. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, **33**, 103−119.
35. Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503−571.
36. Hughes, S.H., Shank, P.R., Spector, D.H., Kung, H.-J., Bishop, J.M., Varmus, H.E., Vogt, P.K. and Breitman, M.L. (1978) *Cell*, **15**, 1397−1410.
37. Muller-Neumann, M., Yoder, J.I. and Starlinger, P. (1984) *EMBO J.*, **198**, 19−24.
38. Pohlman, R.F., Fedoroff, N.V. and Messing, J. (1984) *Cell*, **37**, 635−643.
39. Pohlman, R.F., Fedoroff, N.V. and Messing, J. (1984) *Cell*, **39**, 635−643.
40. Dooner, H.K., English, J. and Ralston, E.J. (1988) *Mol. Gen. Genet.*, **211**, 485−491.
41. Mullis, K.A. and Faloona, F.A. (1987) In Wu, R. (eds), Methods in enzymology. Academic Press, San Diego, CA.,155, 335−350.
42. Gray, J.C., Kung, S.D., G., W.S. and Sheen, S.J. (1974) *Nature*, **252**, 226−227.
43. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.
44. Schwartz, D. and Dennis, E. (1986) *Mol. Gen. Genet.*, **205**, 476−482.
45. Chomet, P.S., Wessler, S. and Dellaporta, S.L. (1987) *EMBO J.*, **6**, 295−302.
46. Tsiapalis, C.N. and Narang, S.A. (1970) *Biochem. Biophys. Res. Comm.*, **39**, 631−636.
47. Sgaramella, V. and Khorana, H.G. (1972) *J. Mol. Biol.*, **72**, 427−444.
48. Wiaderkiewicz, R. and Ruiz-Carillo, A. (1987) *Nucl. Acids Res.*, **15**, 7831−7848.
49. Wu, D.Y. and Wallace, R.B. (1989) *Gene*, **76**, 245−254.
50. McClintock, B. (1967) *Carnegie Inst. Year Book*, **65**, 568−578.

51. Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Z. and Saedler, H. (1986) *EMBO J.*, **5**, 835–841.
52. Masson, P., Surosky, R., Kinsbury, J.A. and Fedoroff, N.V. (1987) *Genetics*, **117**, 117–137.
53. Banks, J.A., Masson, P. and Fedoroff, N. (1988) *Genes Dev.*, **2**, 1364–1380.
54. Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. (1988) *Genetics*, **118**, 519–526.
55. Seed, B.A., Parker, R.C. and Davidson, N. (1982) *Gene*, **19**, 201–209.
56. Bennet, M.D. and Smith, J.D. (1976) *Phil. Trans. Royal Soc. Lond.*, **274**, 227–274.
57. Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. (1984) *Mol. Gen. Genet.*, **194**, 15–23.
58. Silver, S. and Keerikatte, V. (1989) *J. Virol*, **63**, 1924–1928.