

Steroid hormone responsiveness of a family of closely related mouse proviral elements

Charu Ramakrishnan, Diane M. Robins

Department of Human Genetics, University of Michigan, Medical Science II M4708, Ann Arbor, Michigan 48109-0618, USA

Received: 29 April 1997 / Accepted: 14 July 1997

Abstract. Regulation of the mouse sex-limited protein (*Slp*) gene is unusual in that hormone response is conferred by the 5' LTR of an upstream inserted provirus, dubbed the imposon (*imp1*). In a search for additional genes whose regulation has been affected by retrotransposition events, we isolated two partial proviral elements by stringent screening of a mouse genomic library. One clone (*imp2*) contained a portion of the envelope gene and a 3' LTR that was nearly identical to the 3' LTR of *imp1*; this similarity extended to insertion into a B1 repetitive element. The second proviral clone (*imp3*) contained a 5' LTR and associated coding sequences, but lacked its 3' LTR; the LTR of *imp3* differed by 12% from the *imp1* sequence. To assess potential hormone response, proviral enhancer regions cloned into reporter vectors were tested in transfection. The *imp2* enhancer was similar in behavior to *imp1*, conferring both androgen and glucocorticoid induction in one fragment context and an androgen-specific response in another. In contrast, the *imp3* enhancer allowed high expression in the absence of hormone and was less responsive to steroids in general and androgen in particular. These three proviral elements define a small family of steroid responsive proviruses in the mouse genome, and at least one member has had a lasting impact on an endogenous gene's regulation.

Introduction

Retroviral-like elements have had a significant impact on the eukaryotic genome, as demonstrated by the vast number and variety of repetitive chromosomal sequences derived from the conversion of RNA to DNA. In addition to a large effect on overall genome structure owing to this bulk of repetitive DNA, retrovirus-like elements themselves provide a rich source of genetic variation and potential for mutation of host genes. Examples of their effects underlie phenomena as diverse as color variegation in maize (McClintock 1956), hybrid dysgenesis in *Drosophila* (Bingham et al. 1982), and several instances of mammalian oncogenesis (Hayward et al. 1981).

In both human and mouse genomes, endogenous proviruses and retrovirus-like elements number several thousand (Coffin 1984; Larsson et al. 1989). Integration of these elements often disrupts endogenous gene expression, such as in the mouse dilute and hairless loci that are inactivated by Moloney murine leukemia virus (MuLV; Jenkins et al. 1981; Stoye and Coffin 1988). Other insertions lead to gene activation, such as the classic example of *c-myc* activation in ALV-induced lymphoma (Hayward et al. 1981). While many examples of effects on gene expression have been noted, few cellular genes have been found stably associated with retrovirus-like elements that influence their expression. This may be because most such insertions are deleterious or inherently

unstable, and thus lost from the population. However, it is also possible that such insertion events have gone unrecognized because they occurred long ago—remnants of earlier active particles may now be requisite components of wild-type gene expression. The large number of such elements in the genome, whose antiquity is seen in their significant divergence from active proviruses (Dunwiddie et al. 1986), fosters speculation on their role in evolution of endogenous gene expression (Robins and Samuelson 1992; Britten 1996).

In the course of studying androgen regulation of the mouse sex-limited protein (*Slp*), a variant but functional complement component (Van Den Berg et al. 1992), we encountered a novel instance in which an ancient proviral insertion has conferred hormonal dependence on a cellular gene (Stavenhagen and Robins 1988). An androgen-responsive enhancer of *Slp* was localized 2 kb upstream of the transcription start site by several criteria: position of male-specific DNase I hypersensitive sites (Hemenway and Robins 1987), hormone-dependent activity of a DNA fragment from this region in transfection assays (Lorenz et al. 1988), and abrupt sequence divergence of this region from the corresponding flank of the homologous but not hormonally regulated adjacent *C4* gene (Stavenhagen and Robins 1988). The enhancer proved to reside within the 5' LTR of an oppositely oriented provirus, whose full-length reveals vestiges of a C-type retrovirus, most similar to the MuRRS (murine retroviral-related sequences) family (Schmidt et al. 1985). Sequence data, including divergence of the two LTRs, suggest that this element inserted millions of years ago and is linked to *Slp* in all lab mouse strains (Stavenhagen and Robins 1988).

Given the effect of the *Slp* provirus in altering gene regulation in an evolutionarily stable manner, we were interested in the generality of this event. Therefore, we sought closely related proviruses that might similarly impose androgen-dependent regulation on neighboring genes. In a stringent screen, two additional retrovirus-like sequences were isolated from a mouse genomic library. One element is highly similar to the *Slp* provirus, even in so far as its insertion into a B1 element. While its enhancer is somewhat weaker than that of the *Slp* provirus, it retains preference for induction by androgen. The LTR of the other element isolated is more divergent in sequence; the enhancer region of this element confers much higher basal transcription in transfection, with less hormonal response above this level. These elements provide an opportunity to find evidence of further influences on the mouse genome, and to ask whether their regulation is evolving towards or away from androgen-dependence.

Materials and methods

DNA probes. Proviral DNA fragments for genomic library screening and Southern blot hybridizations were as follows. The LTR probe was a 500-bp *DdeI* fragment encompassing most of the 5' LTR of the *Slp* provirus, extending into the *gag* 5' untranslated region. The enhancer region probe

was the 160-bp C' Δ 2 fragment excised from a pGem3 subclone as a 180-bp *Bam*HI-*Sac*I fragment. A 250-bp *Xho*II fragment within the *envelope* region of the provirus was used as the *env* probe. All fragments were electroeluted from agarose gels and labeled with α^{32} P-dCTP by random priming with hexamers.

Southern blotting. Mouse genomic or phage DNAs were digested with 2–5 U restriction enzyme per μ g DNA, fractionated by electrophoresis through agarose gels, and transferred to Gene Screen Plus (NEN) or Zetaprobe (Biorad) membranes. Mouse genomic blots were hybridized at 68°C for 12–16 h in 0.25 M NaHPO₄ (pH 7.5), 7.0% SDS, and washed to final stringencies of 0.5, 0.2, or 0.1 \times SSC, 0.1% SDS, at 72°C, as indicated in figure legends. Phage blots from which restriction maps were constructed were hybridized at 65°C for 2 h in RapidHyb solution (Amersham) and washed at 72°C in 0.1 \times SSC, 0.1% SDS.

Library screening. A genomic library in λ EMBL3 was constructed from B10.BR mouse liver DNA partially digested with *Sau*3A. For screening, approximately 8×10^5 plaques were immobilized on nitrocellulose filters (Gelman Sciences) as per manufacturer's instructions. The filters were hybridized with the three proviral probes (LTR, C' Δ 2, *env*) or *Slp* cDNA in 4 \times SET (1 \times SET is 0.3 M NaCl, 60 mM Tris-HCl, pH 8.0, 4 mM EDTA, pH 8.0), 5 \times Denhardt's, 0.1% SDS, and 1 \times NaPi/PPi at 68°C overnight. Washing was at 70°C in 0.1 \times SSC, 0.1% SDS. Overnight exposures to Kodak film at –70°C were sufficient for detection. Plaques that were positive with the three proviral probes but not *Slp* cDNA were purified for further analysis.

Phage characterization. Phage DNA was prepared as described (Maniatis et al. 1982). 2 μ g of DNA from each phage was digested with *Bam*HI or *Bgl*II, followed by *Sal*I, separated on a 0.7% agarose gel, and transferred to Zetaprobe membrane with a Hoefer Scientific vacublotter. Blots were baked at 80°C under vacuum, probed with C' Δ 2 at 65°C for 2 h in RapidHyb solution (Amersham), and washed as before. For each phage, the *Bam*HI or *Bgl*II fragment with homology to C' Δ 2 was subcloned into pGem4 (Promega). These and other phage subclones were sequenced with Sequenase, version 2 (USB), with either commercial T7 and SP6 primers or primers synthesized complementary to regions of the inserts based on obtained sequence. Sequence comparisons were made with the NIH, NCBI Blast system (Altschul et al. 1990).

Reporter constructs and transfection. Construction of C' Δ 2tkCAT and C' Δ 9tkCAT has been previously described (Adler et al. 1991). The fragment equivalent to C' Δ 2 from *imp2* was obtained as a *Apa*I-*Hin*II fragment, like C' Δ 2 itself, and inserted into the pGem3-tkCAT vector (Stavenhagen et al. 1987) at the *Sma*I site. The equivalent fragment from the 3' LTR of *imp1* was similarly cloned. The single base difference of *imp2* within the HRE-3 sequence was introduced into C' Δ 2, C' Δ 9 and other constructs by PCR-mutagenesis with the T7 primer and the HRE-3 mutant oligonucleotide (HRE-3') TAGAGGATCCCCCTGAAA TAGCCTG. Resulting fragments were digested with *Bam*HI and *Sac*I and inserted into the similarly cleaved tkCAT vector. To examine separately the functional effects of differences between the four *imp* LTRs, fusions were created via a *Mam*I site in the enhancer (at base 175 of the LTR, between the 12-b insertion and the 14-b deletion of *imp2*, and conserved in all four LTRs) and the *Bgl*II site in the ampicillin resistance gene of the plasmid (see Fig. 7 for diagrammatic representation of the recombinants). An *Apa*I-*Hin*II fragment of *imp3*, subcloned into *Sma*I-cleaved tkCAT, generated a C' Δ 9-like construct. A C' Δ 2 equivalent from *imp3* was obtained from this latter construct by joining the *Mam*I-*Bgl*II fragment containing the 3' region of the enhancer through the CAT gene, to the *Mam*I-*Bgl*II fragment of C' Δ 2 containing the 5' end of the enhancer. All reconstructed enhancers were sequenced for correctness.

Receptor expression vectors used in these experiments were mouse AR and rat GR, originally obtained from D. Tindall and K. Yamamoto, respectively, and subcloned into pCMV5 (Andersson et al. 1989). The maintenance of CV-1 cells, transfection by DEAE-dextran, and performance of CAT assays were all as described previously (Adler et al. 1993). Briefly, transfections contained 6 μ g reporter plasmid, 300 ng receptor plasmid, and receptors were activated with 10^{-7} M dihydrotestosterone for AR or 10^{-6} M dexamethasone for GR. All data points are derived from three to five independent transfections.

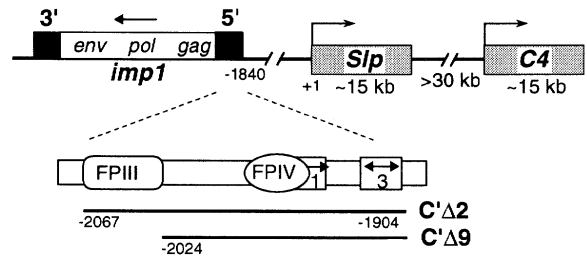


Fig. 1. The *Slp* locus. The provirus upstream of *Slp* is shown as arranged within the mouse major histocompatibility complex (not to scale). *Slp* and *C4* are tandemly duplicated genes, more than 95% identical in nearly 20 kb of sequence (Ogata and Zepf 1991). The provirus, nicknamed *imp1* (for *imposon1*), is oriented opposite to the *Slp* direction of transcription. The hormonally responsive enhancer region characterized within the 5' LTR is diagrammed below, showing several nonreceptor factor binding sites (FPIII, FPIV), as well as the hormone response elements HRE-1 (a half-site designated by an arrow) and HRE-3 (a palindromic consensus element indicated by the double-headed arrow). The 160-bp enhancer fragment C' Δ 2 responds to glucocorticoid as well as androgen receptor, whereas the 120-bp subfragment C' Δ 9 specifically responds only to androgen, as does the *Slp* gene *in vivo*.

Results

The provirus upstream of the mouse *Slp* gene is unusual in two regards. Firstly, it appears to have been maintained after chromosomal integration over evolutionary time, because the association with *Slp* is evident in all inbred lines of mice, and the sequences of the 5' and 3' LTRs, which should be identical following retroviral replication, diverge by at least 5% (Stavenhagen and Robins 1988). Secondly, the provirus appears to have imposed hormonal regulation on *Slp* expression, because it is found associated with androgen-dependent, but not androgen-independent, *Slp* alleles (Stavenhagen et al. 1987), and an androgen-responsive transcriptional enhancer has been defined within its 5' LTR (Adler et al. 1991). A diagram of the provirus, *imp1*, in relation to *Slp* and its homologous neighbor *C4* is shown in Fig. 1. C' Δ 2 is a fragment from the proviral enhancer region that confers steroid responsiveness on heterologous reporter genes in transfection; the C' Δ 9 subfragment is selective in response only to androgen (Adler et al. 1992).

To demonstrate further the antiquity of this proviral insertion, we examined Southern blots of DNA from several wild-derived mice (Fig. 2). C' Δ 2 was used as a probe to reveal a subset of potentially hormone-responsive elements. The right lane of Fig. 2 contains B10.D2 DNA, from which we first characterized genomic clones of this region; arrows indicate the 2.9-kb and 3.8-kb *Bam*HI fragments that encompass the 5' and 3' LTRs, respectively, of *imp1*. The 2.9-kb fragment has one terminus within the 5' untranslated region of *Slp* and the other in the *gag* region of the provirus, with 1.8 kb of *Slp* 5' flank in between (Stavenhagen et al. 1987). These two bands, and additional hybridizing fragments, appeared similar in DNA from two *t* locus strains. These strains were of interest because their Chr 17, where *Slp* resides within the major histocompatibility complex, is nonrecombinogenic owing to several large inversions and may represent a more ancient *Mus* chromosome (Hammer et al. 1989). Commensal mouse strains (*Mus musculus*, *M. molossinus*, and *M. castaneus*) also showed the 2.9- and 3.8-kb bands suggestive of *imp1* association with *Slp*; however, there were differences in other fragments that cross-hybridized. The more distantly related aboriginal species, *M. hortulanus* and *M. spretus*, also shared the *imp1*-specific 2.9-kb fragment but diverged in hybridization pattern elsewhere. Even the mice most distantly related to *M. domesticus* in this panel, *cervicolor* and *cooki*, evidenced what may be *Slp* proviral bands, but differed substantially in other hybridizing fragments. In sum, this corroborates DNA sequence analysis suggesting that the provirus

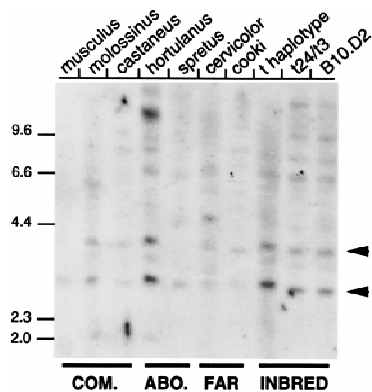


Fig. 2. Presence of *imp1*-related proviruses in wild-derived and divergent mouse species. 5 μ g of DNA of each of the mus species or strains were restricted with *Bam*HI, electrophoresed, and transferred to nitrocellulose as in Methods. The mice are grouped by commensal species (COM.), aboriginal (ABO.), distantly-related (FAR), or inbred lines. The blot was probed with the C' Δ 2 fragment of the *imp1* LTR and washed to a final stringency of $0.2 \times$ SSC at 72°C. The arrowheads indicate the 2.9- and 3.8-kb *Bam*HI fragments that encompass the 5' and 3' LTRs, respectively, of *imp1*.

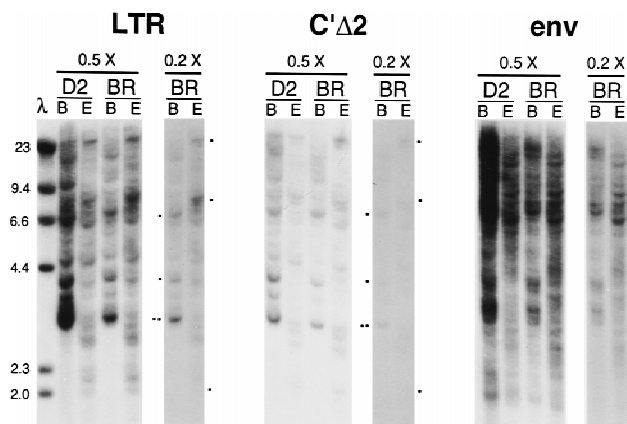


Fig. 3. Numerous genomic sequences similar to *imp1* suggest a family of related proviruses. B10.D2 and B10.BR mouse DNAs (10 μ g each) were restricted with *Bam*HI or *Eco*RI, electrophoresed, and transferred to nitrocellulose. Separate filters were probed with a 500-bp fragment encompassing the 5' LTR, a 160-bp fragment of the enhancer C' Δ 2, or a 250-bp fragment from the *env* region, and were washed with either $0.5 \times$ SSC or $0.2 \times$ SSC, 72°C, as indicated. Dots around the filter of B10.BR DNA probed at high stringency with C' Δ 2 indicate fragments containing *imp* LTRs described in the text, and are similarly marked on the panel probed with the LTR where these bands have greater intensity among other hybridizing fragments. Double dots indicate a doublet containing the 2.9- and 3.0-kb *Bam*HI fragments of *imp1* and *imp3* (see text).

upstream of *Slp* inserted several million years ago in mouse evolution. The constancy of the 2.9- and 3.8-kb fragments relative to other hybridizing fragments suggests that the arrangement of this chromosomal region may be conserved for some reason.

Stringent hybridization with the C' Δ 2 enhancer fragment of *imp1* revealed a limited number of cross-hybridizing elements. This was in contrast to hybridization with the entire LTR, which even at high stringency hybridized to many bands (Fig. 3). We hypothesized that sequences most similar to C' Δ 2 represent hormone-responsive, proviral LTRs that might be androgen specific in their response. To isolate such elements, we screened a genomic library from the B10.BR strain, which is congenic to B10.D2, differing only within the MHC region. No significant differences in the hybridization pattern of these two strains were evident with three proviral probes (Fig. 3), and a discrete subset of bands was

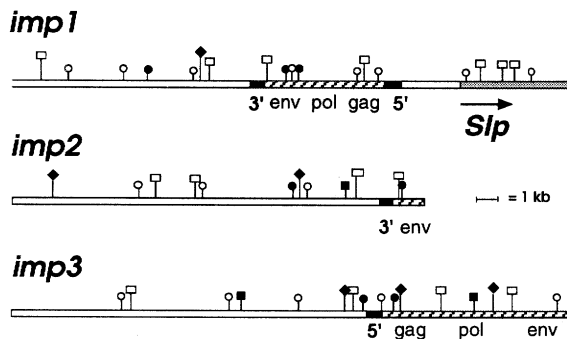


Fig. 4. Three similar proviral elements within the mouse genome. A composite restriction map of *imp1* (Stavenhagen and Robins 1988) and the 5' end of *Slp* (Hemenway and Robins 1987) are shown in comparison with maps of the phage inserts containing partial copies of *imp2* and *imp3*. Restriction sites are shown for *Bam*HI (\circ), *Bgl*III (\square), *Eco*RI (\blacklozenge), *Hin*DIII (\bullet), and *Pst*I (\blacksquare).

apparent with the C' Δ 2 probe. The library was first screened for clones that hybridized to a 500-bp fragment encompassing the *imp1* 5' LTR. These phage were then rescreened with an *env* fragment to insure obtaining retroviral-like elements, and at high stringency with C' Δ 2 to isolate elements with greatest similarity to the hormone-responsive enhancer. The plaques were also counter-screened with *Slp* cDNA to identify re-isolates of *imp1*.

Two phage meeting these criteria, dubbed *imp2* and *imp3*, were analyzed further; restriction maps of their respective 16- and 23-kb phage inserts are compared in Fig. 4 with the original *imp1* and *Slp*. Southern blotting and partial sequence analysis of the inserts showed that each phage contained a single LTR and various extents of proviral coding regions. No matches to known genes in the database were found on the basis of preliminary analysis of non-viral regions of the subcloned inserts. These LTRs accounted for prominent bands that hybridized to C' Δ 2 at high stringency (Fig. 3). The *imp2* LTR is encompassed by a 6.6-kb *Bam*HI and an 8-kb *Eco*RI fragment, the *imp3* LTR is within a 3-kb *Bam*HI and a 2-kb *Eco*RI fragment, and the *imp1* LTRs, as described before, account for the 2.9- and 3.8-kb *Bam*HI fragments, with the entire *imp1* and *Slp* gene contained within a 30-kb *Eco*RI fragment. Thus, from the Southern blot data, we appear to have isolated the two proviruses in the genome most similar to the *imp1*.

The degree of similarity was assessed by sequence comparison of the LTRs of *imp1*, 2, and 3 (Fig. 5). As noted previously (Stavenhagen and Robins 1988), the 5' and 3' LTRs of *imp1* differ from each other by 5% in single base changes, by a 12-bp addition in the 3' LTR relative to the 5' LTR, and by 9 bp in the 5' LTR that are not present in the 3' LTR. Intriguingly, the 3' LTR of *imp2* is nearly identical to the 3' LTR of *imp1*, differing only in one base (position 78 of the LTR) and lacking 14 bp more centrally. The 5' LTR of *imp3* shows more extensive scattered base changes, about 12% relative to *imp1*, prior to extensive loss of similarity at the 3' end of the LTR. The *imp3* LTR is similar to *imp1*'s 5' LTR in lacking the 12-bp insertion of both *imp1* and *imp2* 3' LTRs, and contains the 9 bp missing from the two 3' LTRs. The differences of these LTRs within the C' Δ 2 enhancer region are diagrammed in Fig. 5B.

In view of the surprising similarity of the 3' LTRs of *imp1* and *imp2*, which is greater than the similarity of *imp1*'s 3' LTR to its own 5' LTR, these elements were compared further with respect to internal proviral coding regions and to chromosomal sequences at the insertion sites (Fig. 6). A 180-bp sequence of *imp2* from the end of the phage insert aligned with the 3' end of the *pol* gene of *imp1* (Fig. 6A) and showed significant base difference (nearly 20%), confirming that these elements are distinct and not nearly identical recent insertions. Both elements inserted into B1 repeti-

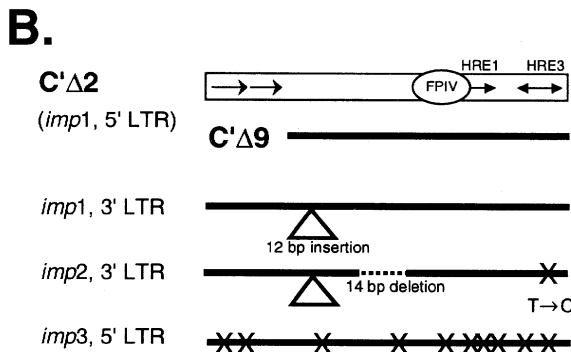
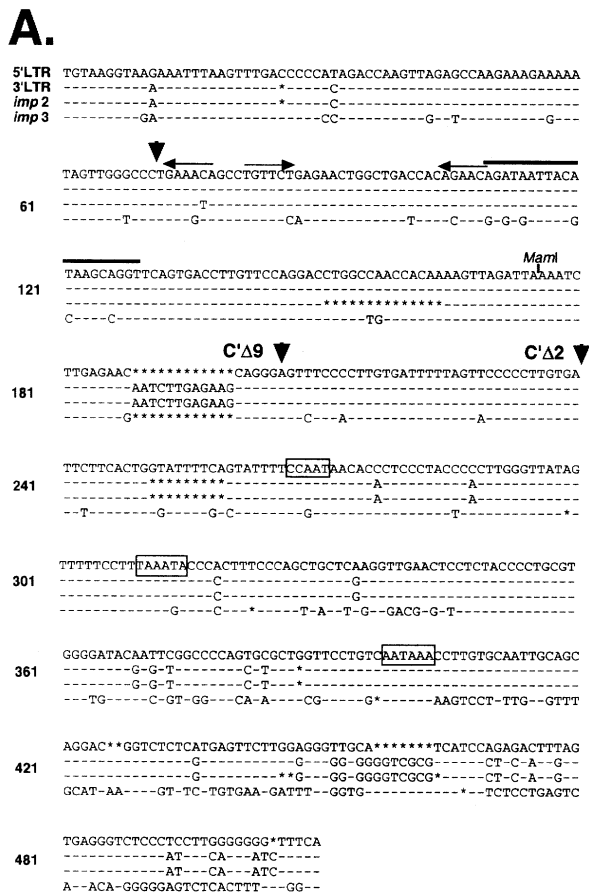


Fig. 5. Comparison of LTR sequences and enhancer regions. (A) The DNA sequences of the LTRs of *imp2* and *imp3* aligned with the 5' and 3' LTRs of *imp1*. Numbering is shown for the *imp1* 5' LTR, with 1 being the first base proximal to *Slp*. Dashes indicate identity, and asterisks denote missing bases. Large arrowheads indicate the limits of the C'Δ2 and C'Δ9 enhancer fragments, horizontal arrows indicate HREs, and a bar indicates the FPIV binding site. Boxes mark conserved CAAT, TATA, polyadenylation elements. (B) Schematic comparison of the enhancer fragments. Functional elements within the 160-bp C'Δ2 fragment of the *imp1* 5' LTR are shown, with the 40-bp shorter C'Δ9 aligned below. These fragments are drawn in their orientation to the *Slp* gene, that is, inverted with respect to the LTR sequence above. The *imp1* 3' LTR has a 12-bp insertion relative to the 5' LTR; the *imp2* 3' LTR has this same insertion, a 14 bp deletion and a single base change within HRE-3. The *imp3* 5' LTR has scattered base changes throughout and lacks the insertion found in the two 3' LTRs.

tive elements (Fig. 6B), with some further sequence similarity in the chromosomal sites prior to abrupt divergence. The B1 sequences themselves are not identical, and it is not without precedent for retrotransposons to insert within other transposed ele-

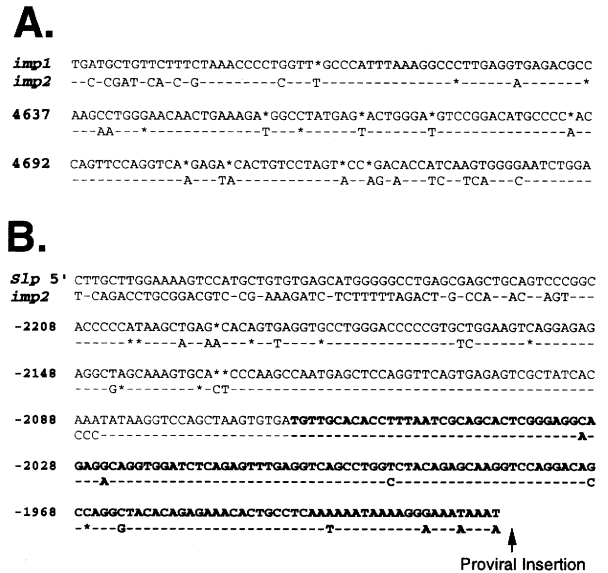


Fig. 6. Comparison of *imp1* and *imp2* proviral and flanking sequences. (A) Bp 4578–4748 (with respect to the first base of the 5' LTR) of *imp1* (top line) are aligned with the similar region of *imp2* below. Dashes indicate identity, asterisks are missing bases, and different bases are written in. (B) *Slp* 5' flank is shown, numbered similarly to *C4* 5' flank (that is, excluding the 6027 bp of *imp1*), to the point of insertion of *imp1*. The chromosomal sequence adjacent to the 3' LTR of *imp2* is aligned to reveal similarity of the insertion into a B1 repetitive element (B1 sequences are bold), with further sequence similarity beyond, prior to complete divergence.

ments (Robins and Samuelson 1992). Further examination of these two genomic regions may unravel what appears to be a nonhomologous recombination event(s) subsequent to proviral integration (see Discussion).

Since *imp1*'s enhancer confers androgen dependence on *Slp* expression (Adler et al. 1991), we asked whether similar DNA sequences of *imp2* and *imp3* had the capacity to confer hormonal response on heterologous genes in transient transfection assays. Analysis of the *imp1* 5' LTR has revealed that several elements are involved in hormonal response, including a near-consensus hormone response element (HRE-3), an HRE half-site (HRE-1), and multiple DNA binding sites for nonreceptor transcription factors (Fig. 1). The nonreceptor factors, rather than the receptor binding site itself, determine the hormonal specificity of the enhancer: in the context of C'Δ9, the enhancer can be activated by androgen but not glucocorticoid receptor (AR, GR), whereas C'Δ2 can respond to both receptors (Adler et al. 1992). The region of C'Δ2 lacking from C'Δ9 (FPIII) confers a high basal (uninduced) expression and allows activation by GR as well as AR (Scheller et al. 1996). Of the nonreceptor binding sites within the androgen-specific C'Δ9 subfragment, FPIV appears crucial on the basis of *in vivo* footprinting data (Scarlett and Robins 1995).

Fragments equivalent to C'Δ2 and C'Δ9 were obtained from *imp1* and *imp2*'s 3' LTR and *imp3*'s 5' LTR with conserved restriction enzyme sites, or were created by fusing portions of the other LTRs with C'Δ2 or C'Δ9 at the conveniently located *MamlI* site (see Methods). The single-base T-to-C difference within the HRE-3 of *imp2* was introduced in some plasmids with a mutagenic oligonucleotide. This set of enhancers driving the tkCAT reporter gene was transiently transfected into CV-1 cells, along with expression vectors encoding AR or GR, and incubated with or without the appropriate hormone for 44 h. CAT activity in cell lysates was determined as percentage conversion of input chloramphenicol to the acetylated form, for three independent transfections. The histogram of Fig. 7 shows CAT levels in the presence of hormone; basal activity of the plasmid in the absence of hormone and fold

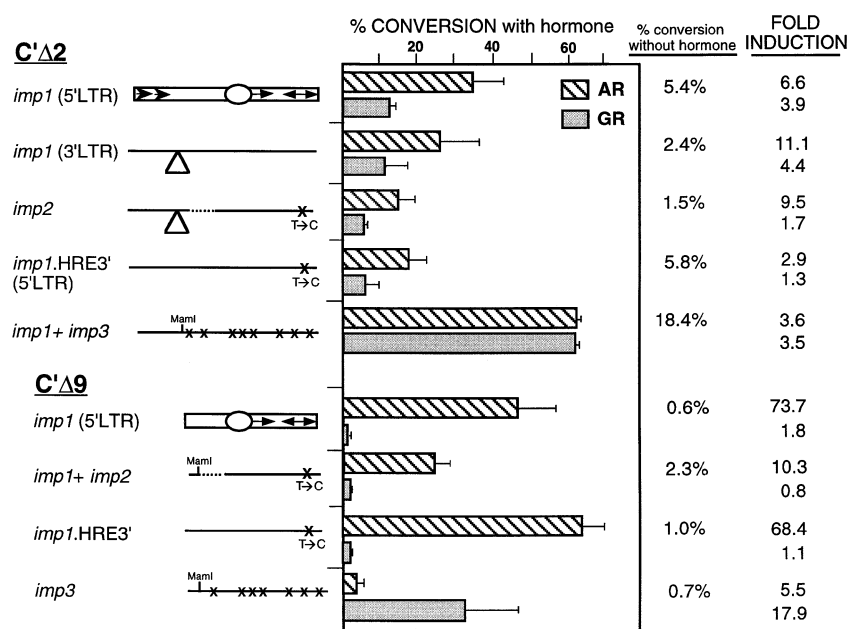


Fig. 7. Hormonal response of LTR fragments. The 160-bp C'Δ2 fragment of *imp1*'s 5' LTR has a moderate level of expression in the absence of hormone that is inducible by either androgen or glucocorticoid. In contrast, the 120-bp C'Δ9 fragment has undetectable basal expression and responds strongly and specifically to androgen. Comparable fragments of the other *imp* LTRs cloned before the tkCAT reporter gene were tested in transient transfection into CV-1 cells with receptor expression plasmids (AR, GR). Constructs are depicted diagrammatically on the left and their construction described in Methods. The triangle indicates the 12-bp insertion, the dotted line a 14-bp deletion, the T-to-C shows the single base change in HRE-3, and the Xs over *imp3* sequences connote significant sequence divergence. Induced levels of CAT activity and standard error of the mean are shown as percentage chloramphenicol conversion for three independent assays (S.E.M. indicated), with hatched bar indicating dihydrotestosterone-induced level and gray bar showing dexamethasone-induced level. Expression in the absence of hormone and fold induction are shown in columns to the right; the inductions for C'Δ9-like fragments are generally higher than for C'Δ2-like fragments owing to lower basal expression of C'Δ9.

induction (activity with hormone relative to without) are tabulated to the right.

Most C'Δ2-like enhancers (the upper four shown in Fig. 7) showed expression in the absence of hormone significantly greater than that of the tkCAT reporter without an enhancer (about 0.6% conversion under these conditions). These enhancers were moderately inducible, with androgen eliciting about a two-fold greater response than glucocorticoid. The 3' LTR of *imp1*, differing solely by the 12-bp insertion, had lower uninduced expression than the 5' LTR, but attained similar induced levels of expression. The equivalent *imp2* fragment had lower basal expression and was induced by androgen nearly tenfold, but was less than twofold responsive to glucocorticoid. Interestingly, when the T-to-C base change within HRE-3 was placed in the context of the original C'Δ2 (*imp1.HRE3'*), the effect was to decrease induction by both hormones. In sharp contrast to *imp1* and *imp2*, the C'Δ2-like fragment of *imp3* (constructed with *imp1* sequences at the 5' end by fusion at the *MamI* site) had a very high level of expression in the absence of hormone and was induced about 3.5-fold above that by either steroid.

By comparison with C'Δ2, C'Δ9 has negligible basal expression and is exquisitely sensitive to androgen but not glucocorticoid. Introduction of *imp2*'s single base change in HRE-3 into the context of C'Δ9 had little significant effect, whereas conjoining this with the 14-bp deletion increased the uninduced level of expression and reduced maximal activity obtainable with androgen. However, neither of these differences altered the fundamentally androgen-specific response of this fragment. *Imp3*, on the other hand, showed markedly better induction by glucocorticoid than by androgen as the truncated C'Δ9 fragment, with low basal expression as anticipated from the deletion of the C'Δ2-specific 5' end. Integrating these data suggests that *imp2*, like *imp1*, is a weak enhancer in the absence of steroid, and shows a strong preferential induction by androgen. *Imp3*, on the other hand, is a strong general enhancer, moderately inducible by steroid, with no preference for androgen induction.

Discussion

The provirus upstream of the *Slp* gene, *imp1*, provides an intriguing example of a mammalian retrotransposon that has become

heritably associated with cellular gene regulation (Stavenhagen and Robins 1988). New patterns of expression for the rat oncomodulin gene (Banville and Boie 1989), a placenta-specific mouse gene (Chang-Yeh et al. 1991), and the human amylase gene (Ting et al. 1992) are also now correlated with insertion of retrovirus-like elements. Although these known instances are few in number, the phenomenon may be more widespread, since it is estimated that more than 10% of the mammalian genome originated from reverse transcribed sequences (Nouvel 1994). Further, as many as 5% of recessive mutations in inbred mice may be due to proviral insertions (Stoye and Coffin 1988). We have explored the importance of retrotransposons in emergence of new regulatory networks in evolution by first isolating elements similar to *imp1* and then assessing their potential for genomic impact by their hormonal induction of a reporter gene in vitro.

The usefulness of *imp1* to probe for additional elements influencing genomic expression and evolution resides in both its age and the precise type of regulation conferred. *Imp1*'s antiquity demonstrates at least some capacity for chromosomal stability, whereas more recent integration events may not be ultimately heritable. That the chromosomal insertion of *imp1* occurred several million years ago was first suggested by sequence variation between its LTRs (Stavenhagen and Robins 1988) and is supported here by hybridization to DNA from divergent mouse species. The occurrence in feral mice of a fragment similar in size to the one in lab strains that spans the *imp1-Slp* intervening distance suggests that this gene arrangement existed prior to the separation of commensal and aboriginal species, estimated at 3–6 million years ago (Ferris et al. 1983). Similarly, ancient insertions that have acquired integral functions in higher genomes include a novel repetitive element that is part of the coding sequence of human transaldolase (Banki et al. 1994) and *copia*-like elements found in the flanking regions of plant genes (White et al. 1994). These insertions may differ in some way from the majority of currently active elements, such as *copia* in *Drosophila* or Moloney leukemia virus in mice, that cause many spontaneous mutations but are not found stably associated with endogenous gene expression (Bingham and Zachar 1989).

The two additional elements isolated in this screen account for prominent restriction fragments on high-stringency genomic Southern blots. It is likely that unassigned bands on the blot rep-

resent the additional LTRs of *imp2* and *imp3* that are not contained in the phase inserts. Thus, it seems that the criteria of this screen isolated the two closest relatives of *imp1* and define a small proviral family. The origin of these elements, their relationship to each other, and the precise time of their chromosomal insertion cannot be determined without further sequence and phylogenetic comparisons, and particularly without obtaining the missing LTRs. Intriguing paradoxes are apparent, however. Most particularly, the 3' LTR of *imp2* is more similar to the 3' LTR of *imp1* than the *imp1* 3' LTR is to its own 5' LTR. The sequence similarity extends past the B1 element insertion site, but then abruptly breaks, and the *pol* regions of *imp1* and *imp2* diverge significantly. This suggests that a nonhomologous recombination event (perhaps gene conversion) encompassing the 3' LTRs occurred at the chromosomal level subsequent to integration. Recombination may have been accelerated by homology of B1 elements, just as many mutations are caused by illegitimate recombination between the analogous human Alu sequences (Deininger 1989). In contrast, the scattered 12% base difference between the *imp3* and *imp1* LTRs suggests a divergence time of several million years and supports the idea that these elements are vestiges of an ancient retroviral infection. Isolation of additional flanking sequences, as well as the missing LTRs, may illuminate the complex origin and relationship of members of the *imp* family.

Imp1's hormone responsiveness, particularly its androgen specificity, is striking and provides an assay for functional as well as sequence relatedness of additional elements. It is not unusual for retroviruses to respond to steroid hormones; mouse mammary tumor virus and Moloney murine sarcoma virus have both been used as experimental models for steroid action (for example, Payvar et al. 1983; Miksicek et al. 1986). Sophisticated regulatory responses may allow virus-like elements to take advantage of growth stimulation of their host cells, as perhaps demonstrated by growth factor responsiveness of VL30s (Schiff et al. 1991) or homeobox binding sites in the *Drosophila copia* element (Cavarec and Heidmann 1993). The imposition of steroid responsiveness by proviral insertion might reasonably account for several instances of sexually dimorphic liver and kidney gene expression in rodents, which is distinct from the regulation of these same genes in humans. An experimental advantage of finding androgen response capabilities in an LTR is that the sequence requirements may be more readily defined than for chromosomal *cis*-acting elements that may be dispersed over large distances. This has made the proviral enhancer of *imp1* particularly useful in studying androgen-specific response, which has not been readily unraveled by analysis of cellular enhancers.

The comparison of the structure and function of these enhancers allows clearer definition of DNA sequences required for steroid-specific response, as opposed to those involved primarily in strong basal transcription. The general behavior of the *imp1* and *imp2* LTRs in activation of a reporter gene is qualitatively similar, showing preferential response to androgen over glucocorticoid in the context of the longer C' Δ 2 enhancer fragment, and stringent specificity in response to androgen in the C' Δ 9 context. In contrast, *imp3* exerts strong enhancement of transcription in the absence of hormone that is only somewhat inducible above those levels, with no greater effect of androgen than glucocorticoid. It is significant, therefore, that most of the sequence differences between these enhancers occur in the consensus HRE-3 receptor binding site and in the FPIV region, noted previously for its association with androgen-regulated *Slp* expression in vivo (Scarlett and Robins 1995). The FPIV region is complex, encompassing a weak receptor binding site (HRE-1) and a site that binds several proteins in gel shift assays, including the ubiquitous transcription factor Oct-1 (Scarlett et al. 1997). In *imp3*, the HRE-1 half-site is altered in two of six bases, and the Oct-1 similarity in three of eight bases, relative to *imp1*. These differences are probably sufficient to disrupt protein binding at either or both sites and could account for

the reduced steroid response in general and the loss of preferential activation by androgen receptor. In contrast, the single-base difference in HRE-3 between *imp1* and *imp2* has little effect on differential activation by androgen versus glucocorticoid receptor, and the additional *imp3* difference, based on position, is unlikely to underlie the reduced response of that enhancer (Nordeen et al. 1990). These natural enhancer variants may prove useful in elucidating further the requirements for androgen-specific activation, as they appear to distinguish between reduced specificity and reduced expression in general.

Among this family of elements, so far only *imp1* is associated with expression of an endogenous gene, and whether the particular regulation conferred is an advantage or is simply neutral in evolution is not yet apparent. A possible influence on reproductive behavior of the mouse could arise from greater diversity of secretory proteins in urine. Alternatively, the androgen specificity may advance the selfish goals of the proviral elements, as efficient horizontal and vertical transmission could be accomplished by expression in the testis. Since *imp1* and *imp2* are androgen specific, and *imp3* is not, further investigation may resolve whether these elements are evolving towards, or away from, hormonal specificity. Regardless of the direction, this may clarify the extent of co-evolution of retroviral and host genomes.

Acknowledgments. We thank Mike Hammer for genomic DNAs from *t*-haplotype and feral mice, John Timmer for performing Southern blot analysis of mouse species DNAs while an undergraduate in the lab at Columbia University, and Janine Tompkins for help with phage characterization during a summer at the University of Michigan. Steve Mount, Linda Samuelson, and Alice Telesnitsky provided helpful discussions. This work was supported by National Institutes of Health grant GM 42597.

Note added in proof: GenBank accession numbers for sequences presented are U95781, U95782, U95783. Due to the symbol *Imp1* being used for the gene imprintor 1, impositions 1, 2 and 3 will be referred to as *imps1*, *imps2*, *imps3*, to avoid confusion.

References

- Adler AJ, Scheller A, Hoffman YY, Robins DM (1991) Multiple components of a complex androgen-dependent enhancer. *Mol Endocrinol* 5, 1587–1596
- Adler AJ, Danielsen M, Robins DM (1992) Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors. *Proc Natl Acad Sci USA* 89, 11660–11663
- Adler AJ, Scheller A, Robins DM (1993) The stringency and magnitude of androgen-specific gene activation are combinatorial functions of receptor and nonreceptor binding site sequences. *Mol Cell Biol* 13, 6326–6335
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215, 403–410
- Andersson S, Davis DN, Dahlback H, Jornvall H, Russell DW (1989) Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J Biol Chem* 264, 8222–8229
- Banki K, Halladay D, Perl A (1994) Cloning and expression of the human gene for transaldolase: a novel highly repetitive element constitutes an integral part of the coding sequence. *J Biol Chem* 269, 2847–2851
- Banville D, Boie Y (1989) Retroviral long terminal repeat is the promoter of the gene encoding the tumor-associated calcium-binding protein oncomodulin in the rat. *J Mol Biol* 207, 481–490
- Bingham PM, Zachar Z (1989) Retrotransposons and the FB transposon from *Drosophila melanogaster*. In *Mobile DNA*. DE Berg, MM Howe (eds) (Washington, D.C.: American Society for Microbiology), pp 485–502
- Bingham PM, Kidwell MG, Rubin GM (1982) The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell* 29, 995–1004
- Britten RJ (1996) DNA sequence insertion and evolutionary variation in gene regulation. *Proc Natl Acad Sci USA* 93, 9374–9377

- Cavarec L, Heidmann T (1993) The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. *Nucleic Acids Res* 21, 5041–5049
- Chang-Yeh A, Mold DE, Huang RCC (1991) Identification of a novel murine IAP-promoted placenta expressed gene. *Nucleic Acids Res* 19, 3667–3672
- Coffin JM (1984) Endogenous viruses. In *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*. R Weiss, N Teich, H Varmus, J Coffin (eds) (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory), pp 1109–1204
- Deininger PL (1989) SINES: short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA*. DE Berg, MM Howe (eds) (Washington, DC: American Society for Microbiology), pp 619–636
- Dunwiddie CT, Resnick R, Boyce-Jacino M, Alegre JN, Faras AJ (1986) Molecular cloning and characterization of gag-, pol- and env-related gene sequences in the ev- chicken. *J Virol* 59, 669–675
- Ferris SD, Sage RD, Prager EM, Ritte U, Wilson AC (1983) Mitochondrial DNA evolution in mice. *Genetics* 105, 681–721
- Hammer M, Schimenti J, Silver L (1989) Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes. *Proc Natl Acad Sci USA* 86, 3261–3265
- Hayward WG, Neel BE, Astrin SM (1981) Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukaemia. *Nature* 210, 475–480
- Hemenway C, Robins DM (1987) DNaseI-hypersensitive sites associated with expression and hormonal regulation of mouse C4 and SIp genes. *Proc Natl Acad Sci USA* 84, 4816–4820
- Jenkins NA, Copeland NG, Taylor BA, Lee BK (1981) Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* 293, 370–374
- Larsson E, Kato N, Cato M (1989) Human endogenous proviruses. *Curr Topics Microbiol Immunol* 148, 115–132
- Loreni F, Stavenhagen J, Kalff M, Robins DM (1988) A complex androgen-responsive enhancer resides 2 kilobases upstream of the mouse *Slp* gene. *Mol Cell Biol* 8, 2350–2360
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press)
- McClintock B (1956) Controlling elements and the gene. *Cold Spring Harbor Symp Quant Biol* 21, 197–216
- Miksicek R, Heber A, Schmid W, Danesch U, Posseckert G, Beato M, Schutz G (1986) Glucocorticoid responsiveness of the transcriptional enhancer of Moloney murine sarcoma virus. *Cell* 46, 283–290
- Nordeen SK, Suh BJ, Kuhnel B, Hutchison CA (1990) Structural determinants of a glucocorticoid receptor recognition element. *Mol Endocrinol* 4, 1866–1873
- Nouvel P (1994) The mammalian genome shaping activity of reverse transcriptase. *Genetica* 93, 191–201
- Ogata RT, Zepf NE (1991) The murine *Slp* gene: additional evidence that sex-limited protein has no biologic function. *J Immunol* 147, 2756–2763
- Payvar F, DeFranco D, Firestone GL, Edgar B, Wrangle O, Okret S, Gustafsson J-A, Yamamoto KR (1983) Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. *Cell* 35, 381–392
- Robins DM, Samuelson LC (1992) Retrotransposons and the evolution of mammalian gene expression. *Genetica* 86, 191–201
- Scarlett CO, Robins DM (1995) In vivo footprinting of an androgen-dependent enhancer reveals an accessory element integral to hormonal response. *Mol Endocrinol* 9, 413–423
- Scarlett CO, Scheller A, Thompson E, Robins DM (1997) Involvement of an octamer-like sequence within a crucial region of the androgen-dependent *Slp* enhancer. *DNA Cell Biol* 16, 45–57
- Scheller A, Scheinman RI, Thompson E, Scarlett CO, Robins DM (1996) Contextual dependence of steroid receptor function on an androgen-responsive enhancer. *Mol Cell Endocrinol* 121, 75–86
- Schiff R, Itin A, Keshet E (1991) Transcriptional activation of mouse retrotransposons in vivo: specific expression in steroidogenic cells in response to trophic hormones. *Genes Dev* 5, 521–532
- Schmidt M, Wirth T, Kroger B, Horak I (1985) Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS). *Nucleic Acids Res* 13, 3461–3470
- Stavenhagen JB, Robins DM (1988) An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 55, 247–255
- Stavenhagen JB, Loreni F, Hemenway C, Kalff M, Robins DM (1987) Molecular genetics of androgen-dependent and independent expression of mouse sex-limited protein. *Mol Cell Biol* 7, 1716–1724
- Stoye JP, Coffin JM (1988) Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* 54, 383–391
- Ting C-N, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 6, 1457–1465
- Van Den Berg CW, Demant P, Aerts PC, Van Dijk H (1992) *Slp* is an essential component of an EDTA-resistant activation pathway of mouse complement. *Proc Natl Acad Sci USA* 89, 10711–10715
- White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. *Proc Natl Acad Sci USA* 91, 11792–11796