

# A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *Drosophila virilis*

(mutagenesis/transposable element/DNA mobilization)

V. SH. SCHEINKER\*, E. R. LOZOVSKAYA\*, J. G. BISHOP†, V. G. CORCES†, AND M. B. EVGEN'EV\*

\*Institute of Molecular Biology, Academy of Sciences of the U.S.S.R., 32 Vavilov Street, Moscow 117984, U.S.S.R.; and †Department of Biology, The Johns Hopkins University, 34th and Charles Streets, Baltimore, MD 21218

Communicated by M. M. Green, September 14, 1990

**ABSTRACT** A hybrid dysgenesis syndrome similar to those described in *Drosophila melanogaster* occurs in *Drosophila virilis* when a laboratory stock is crossed to a wild strain collected in the Batumi region of Georgia (U.S.S.R.). Mutations in various loci obtained during these crosses are presumably induced by the insertion of DNA sequences. We have cloned an induced white mutation and characterized the insertion sequence responsible for the mutant phenotype. This sequence is a 10.6-kilobase (kb) transposable element we have named Ulysses. This element is flanked by unusually large 2.1-kb long terminal repeats. Ulysses also contains other landmarks characteristic of the retrotransposon family, such as a tRNA-binding site adjacent to the 5' long terminal repeat and open reading frames encoding putative products with homology to the reverse transcriptase, protease, and integrase domains typical of proteins encoded by vertebrate retroviruses. Some of the mutations obtained do not contain a copy of the Ulysses element at the mutant locus, suggesting that a different transposable element may be responsible for the mutation. Therefore, Ulysses may not be the primary cause of the entire dysgenic syndrome, and its mobilization may be the result of activation by an independent mobile element.

The term "hybrid dysgenesis" was first used by Kidwell *et al.* (1) to describe a series of phenomena, such as high sterility and mutation rate, male recombination, segregation distortion, and chromosome breakage, that appeared in the progeny of crosses between certain *Drosophila melanogaster* strains. The male recombination and mutagenesis phenomena associated with hybrid dysgenesis had been described by Hiraizumi (2, 3) as associated with the presence in the genome of *MR* elements. These *MR* elements then provide a genetic basis for the mutational and mitotic recombinational components of these events and have now been shown functionally equivalent to the *P* transposable element responsible for the P–M hybrid dysgenesis syndrome (4, 5). The phenotypic characteristics associated with hybrid dysgenesis appear in the progeny of crosses between P males, which contain intact *P* elements in the germ line, and M females, in which this transposon is defective or absent. Some of the distinctive dysgenic traits can still be detected when the reciprocal cross is performed but at much lower levels. The *P* element contains two 31-base-pair (bp) inverted repeats and four open reading frames that encode a transposase involved in the mobilization of both intact and defective *P* elements (5).

A second hybrid dysgenesis system designated I–R gives rise to similar characteristics, such as transient female sterility, X-chromosome nondisjunction, and increased mutation frequency, when an inducer male (I) containing *I* factors is crossed to a responsive (R) female lacking these factors.

The genetic determinant responsible for this phenomenon is the *I* transposable element, which differs from the *P* element in that it encodes a protein with sequence similarities to reverse transcriptase (6). Thus, although the dysgenic traits that arise in P–M and I–R crosses are very similar, the nature of the transposable elements involved is very different. Whereas the *P* element encodes transposase and repressor activities (7, 8) that account for the mechanisms of *P* mobilization and directionality of the dysgenic cross, the *I* element is a poly(A)<sup>+</sup> retrotransposon that encodes two putative open reading frames, the second one with homology to the polymerase and RNase H domains of the reverse transcriptase found in vertebrate retroviruses (6). Some of the dysgenic traits have also been observed in systems involving the hobo family of transposable elements, which can promote high rates of chromosomal instability (9).

Some of us have recently described a dysgenic system that takes place in *Drosophila virilis* in unidirectional crosses between two different strains and results in characteristic traits in the F<sub>1</sub> progeny similar to those seen in the P–M and I–R systems in *D. melanogaster* (10, 11). However, contrary to these, hybrid dysgenesis traits in *D. virilis* such as gonadal sterility in males and females, male recombination, chromosomal nondisjunction, transmission ratio distortion, and the appearance of numerous visible mutations, were observed when males from an established laboratory strain were crossed to wild-type females from strains recently caught from the wild in the USSR. Here we present evidence suggesting that the dysgenic syndrome in *D. virilis* is associated, in part, with the movement of a long terminal repeat (LTR)-containing retrotransposon structurally similar to the proviral form of vertebrate retroviruses.

## MATERIALS AND METHODS

Wild-type strain 9 was collected in Batumi (Georgia, U.S.S.R.) in 1975, and wild-type strain 2 was collected in Kutaisi (Georgia, U.S.S.R.) in 1970. Strain 160 was constructed by introducing a chromosome 6 from stock 104 (USA) containing the glossy mutation into an old laboratory stock (strain 149) from Japan that carries recessive markers in all large autosomes. Wild-type strains of *Drosophila texana* and *Drosophila lummei* that belong to the *virilis* group were also used in these studies.

DNA from *D. virilis* was prepared as described (12). Construction of genomic libraries in  $\lambda$ 7.1 and EMBL3, library screening, DNA restriction and labeling, Southern analysis, etc., were carried out by standard procedures (13). DNA was sequenced by the dideoxynucleotide chain-termination method using Sequenase (14). Computer analysis of DNA sequence homologies was done by using the PC/GENE programs from Intelligenetics. *In situ* hybridization to polytene chromosomes was performed essentially as de-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: LTR, long terminal repeat.



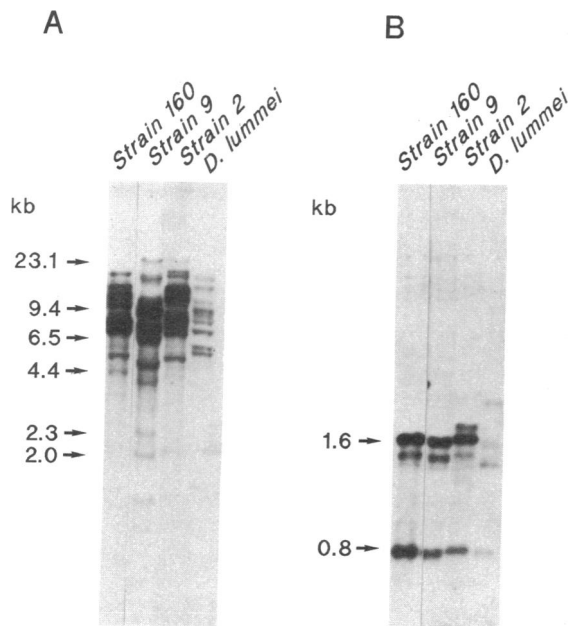


FIG. 3. Southern analysis of Ulysses element. (A) Genomic DNA from strains 160, 9, and 2 of *D. virilis* and wild-type strain of *D. lummei* was digested with *Hind*III, electrophoresed on 1% agarose gel, blotted onto nitrocellulose, and probed with the 2.4-kb *Hind*III-*Eco*RI fragment of Ulysses element labeled with [<sup>32</sup>P]dNTP. (B) Genomic DNA from strains 160, 9, and 2 of *D. virilis* and *D. lummei* was digested with *Hind*III and *Eco*RI and subjected to Southern analysis as described above.

few additional light bands evident in *D. virilis* and, in particular, in *D. lummei* suggest the presence of Ulysses copies with an altered structure.

To gain further insight into the nature of the sequences responsible for the  $w^{d9}$  mutation, we carried out *in situ* hybridizations to polytene chromosomes of the parental strains with the insertion sequence used as probe. Fig. 4A shows the pattern of this hybridization in strain 9. Sequences are present in 15 different euchromatic sites, confirming that the sequences are repeated. In addition, heavy labeling of the centromeric heterochromatin can also be observed. The same type of cytological analysis in strain 160 revealed equivalent intense hybridization at 11 euchromatic sites plus the chromocenter (data not shown). Only 1 site was common in these two strains. Parallel analyses of other *Drosophila* species indicate that Ulysses elements are also present in all species that belong to the *virilis* group, such as *D. texana* and *D. lummei*, although for these, hybridization is concentrated mainly in the chromocenter and only diffuse grains appear in the euchromatin (Fig. 4B). *In situ* hybridization as well as Southern analysis failed to detect similar sequences in *D. melanogaster*, *Drosophila funebris*, and *Drosophila hydei* (data not shown).

To test whether all mutations on the X chromosome obtained from crosses between strains 9 and 160 are caused by insertion of the Ulysses transposon we looked for this element in different mutations by *in situ* hybridization to polytene chromosomes. Ulysses hybridization at the cytological location of the respective loci provides strong evidence for association of this element with the mutation. Results from these experiments indicate that Ulysses insertion took place at the mutant locus in two mutations of white obtained, one of five singed mutations studied, and the only mutation of *Bx* isolated. On the other hand, neither the single yellow nor forked mutations recovered contain Ulysses sequences. We have recently isolated from a yellow and a singed mutation a repeated DNA sequence—in all probability

a transposable element different from Ulysses (M.B.E. and V.G.C., unpublished work). This finding suggests that at least one other transposable element may also be mobilized during hybrid dysgenesis in *D. virilis*.

**Ulysses Element Belongs to the Retrotransposon Family.** Further analysis of the insertion sequences was carried out by determining the DNA structure of the boundary between the Ulysses element and the surrounding white gene in the  $w^{d9}$  mutation and in the normal allele. This analysis showed that the insertion occurred at a cluster of the simple sequence poly[d(G-T-T)]-poly [d(C-A-A)] in the intron region, between the second and third exons of the white gene. This insertion is such that transcription of the element is opposite in direction to that of the white gene (see below). Moreover, the insertion of foreign sequences resulted in a 4-bp duplication in the mutant DNA, a characteristic of transposable elements generated during the integration process (Fig. 2).

The nature of the Ulysses element was determined by analyzing its DNA structure. The element isolated from the  $w^{d9}$  mutation was sequenced and found to be 10.6 kb long with two LTRs 2.1 kb each (Fig. 2). Length of the LTRs is unusual among *Drosophila* transposable elements, which are generally  $\approx 0.5$  kb. The sequence of the central region of Ulysses indicates the existence of additional structural characteristics typical of retrovirus-like transposable elements. For example, a lysine tRNA primer-binding site can be found immediately adjacent to the 5' LTR. In addition, open reading frames that encode putative proteins homologous to vertebrate retroviral products, such as reverse transcriptase, protease, and integrase, can also be found within Ulysses. Fig. 5 shows the amino acid sequence of these regions and a comparison with similar regions of other transposable elements from *Drosophila* and mammals (18–24). The order of these homologies in the Ulysses genome is protease–reverse transcriptase–integrase. This order resembles that found in retroviruses and other retrotransposons, such as gypsy and 17.6 from *D. melanogaster* and *Ty3* from yeast, but different from that in the *Drosophila* copia element and the *Ty1* and *Ty2* elements from yeast (25). These results establish that Ulysses belongs to the family of LTR-containing retrotransposons.

**An Autosomal Inversion Mediated by Ulysses Elements.** The occurrence of rearrangements with breakpoints coincident with mobile elements is a key feature of the dysgenesis systems described so far (26, 27). We carried out extensive polytene-chromosome cytological analysis in search of rearrangements that may be associated with the mutations obtained in dysgenic crosses. We studied all five singed mutations obtained, two white mutations, and the single alleles of miniature, bithorax, yellow, forked, abnormal abdomen, etc., up to a total of 15 mutant stocks. With a single exception, we failed to find any rearrangements in any of the mutants. A stock bearing a singed mutation (located in chromosome 1) was found to have a large heterozygous inversion in chromosome 3 with breakpoints at polytene chromosome sites 32A and 39F, respectively (Fig. 4C). *In situ* analysis of the asynapsed rearranged chromosomes demonstrated copies of Ulysses at the vicinity of both breakpoints (data not shown). Rearrangements of this type may be explained by intrachromosomal pairing and exchange between Ulysses elements located in opposite orientations, as has been demonstrated for the hobo transposable element in *D. melanogaster* (28). It should be emphasized that the finding of this single inversion in *D. virilis* is of importance because spontaneous rearrangements are yet to be reported in the wild populations or laboratory strains of this intensively studied species (29).

## DISCUSSION

We have described a series of genetic phenomena that take place in the progeny of crosses between defined strains of

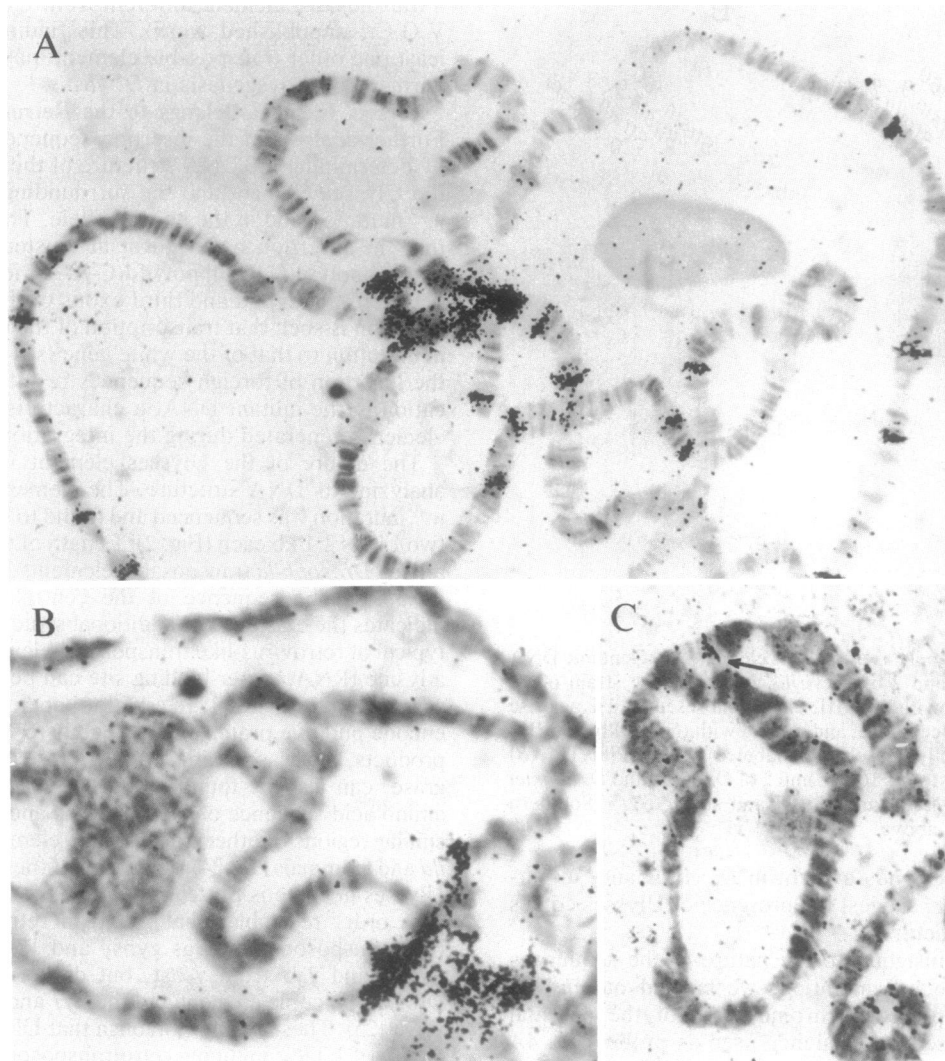


FIG. 4. *In situ* hybridization of  $^3\text{H}$ -labeled Ulysses element to polytene chromosomes from *D. virilis* (A), *D. lummei* (B), and the *sn* strain containing the inversion in chromosome 3 (C). Arrow in C indicates inversion breakpoints labeled by Ulysses. Exposure time was 7 days.

*virilis* (10, 11) and parallel those displayed in the P–M and I–R hybrid dysgenesis syndromes in *D. melanogaster* but differ from them in the nature of the transposable element involved. At least four among ten mutations arising from the *D. virilis* dysgenesis system are caused by insertion of the retrotransposon we have named Ulysses.

Movement of LTR-containing retrotransposons has not been seen during either P–M or I–R hybrid dysgenesis (30). Furthermore, neither transposition bursts nor other known cases of coordinate movements of LTR-containing retrotransposons lead to hybrid dysgenesis in *D. melanogaster* (31, 32). Therefore, the phenomenon described here represents the particular instance of a hybrid dysgenesis syndrome induced or accompanied by mobilization of an LTR-containing retrotransposon.

We have also found a single heterozygous inversion in chromosome 3 in the progeny of a dysgenesis-induced singed mutant. The inversion breaks coincide with Ulysses elements, a situation reminiscent of some *D. melanogaster* dysgenesis systems in which other transposable elements have been found at the inversion breakpoints (25, 26). One of us has previously shown the presence in the *D. virilis* genome of *pDv* elements that occupy  $\approx 200$  sites in the chromosomes. DNA sequence analysis revealed that the *pDv* elements contain tandemly arranged 36-bp repeat units flanked by imperfect direct repeats (11). Moreover, it has also been

shown that the 36-bp sequences in *D. virilis* genome exist in three different orientations relative to one another—i.e., tail-to-tail, tail-to-head, and head-to-head (33). Asymmetrical pairing of *pDv* elements and exchange might create a variety of chromosomal rearrangements that were never observed in nature. It thus appears reasonable to assume that some type of constraints are imposed against pairing and exchanges between intrachromosomal segments in *D. virilis*. These constraints, probably imposed by selection, were overcome by intrachromosomal interactions of Ulysses elements somehow activated by the dysgenic cross that led to the restructuring of chromosome 3 of this *D. virilis* singed stock.

Dysgenesis in *D. virilis* was seen after crossing strains that do not differ significantly in number of Ulysses copies. Thus, one may propose the existence in *D. virilis* of a special copy of the Ulysses element playing the same role as the *MR* element in P–M dysgenesis or the *Mos* factor promoting excisions of the transposable element mariner in *Drosophila mauritiana* (34, 35). On the other hand, the Ulysses element may not be directly responsible for the dysgenesis phenomenon in *D. virilis* because some mutations obtained in the dysgenic crosses did not show hybridization of this element at their normal cytological locations. It is thus formally possible that a second element different from Ulysses directly causes the dysgenic syndrome and that Ulysses mobilization is only a secondary effect of this phenomenon. The move-

A			
ulysses	433	MPFGLC NAAQHFEAHD KVIPANLR SNVVFV	YLDLLIISADFPHTLKYLELVAECLRNANLTIGMAKSKFLFRNLNVLG
412	442	LPFGLK IAPNSFQ RMTIAFSGIEPSQAFI	YMDDLIVIGCSEKHMLKNTLFEVFGKCREYNLKLHPEKCSFFMHVEVTFILG
17.6	334	MPFGLK NAPATFQ RCMNDILRPLLKHKCLV	YLDDIIVFSTSLDEHLQSLGLVFEKLAKANLKLQLDKCEFLKQETTFLG
297	333	MPFGLR NAPATFQ RCMNNILRPLLKHKCLV	YLDIIIFSTSLTEHLNSIQLVFTKLADANLKLQLDKCEFLKKEANFLG
gypsy	309	LPFGLR NASSIFQ RALDDVLRQIGKICYV	YVDDVIFSENESSHVRHIDTVLKLCLIDANMRVSOEKTRFFKESVEYLG
copia	1020	LPQGISCNSD NVCKLNKAIYGLKQAA	YVDDVVIATGDMTRMNNFKRYLMEKFRMTDLNE IK HFIG
MoMLV	308	LPQGFK NSPTLFDEALHRDLAD FRIQ HFDLILLQYVDDLLLAATSELDCCQGTTRALLQTLGNLGYRASAKKAQICQKQVKYL	
HIV-2	333	LPQGWK GSPAIFQHTM RQVLEPFRKA NKDVIIQYMDLIIASDR TD LEHDR	VVLQ LKELLNGLG

B		C		
ulysses	131	AEVEVAGAKMKGLLDTGASVSLIGQG	ulysses 820	H 7aa H 21aa CDTC 118aa SPQSNAAER
412	49	FIHAKTGVKLVFLDGTGADISILKEN	412	901 H 7aa H 29aa CQKC 118aa HQTGVGVVER
17.6	16	ITIKYKNNLKLCLIDTGSTVNMSTKN	17.6	758 H 4aa H 29aa CSIC 110aa KTGVDADIER
297	16	IKIVYKGRSYKCLLDGTGSTINMINEN	297	758 H 4aa H 29aa CNIC 110aa KNGVADVER
gypsy	15	IERRLAGRTLKMLIDTDAAKNYIRPV	gypsy	746 H 3aa H 29aa CRVC 113aa SSSNGQVER
copia	278	VNNTSVMDCNCGFVLDGASDHLINDE	copia	424 H 4aa H 27aa CEC 124aa PQLNGVSE
MoMLV	13	ITLKVGGQPVTFVLVDTGAQHSVLTQN	MoMLV	848 H 3aa H 32aa CKAC 116aa PQSSGQVER
HIV-2	96	VTAYIEGQPVVLLDGTGADDSIVAGI	HIV-2	755 H 3aa H 23aa CAQC 147aa GIGMTPSER

Fig. 5. Comparison of amino acid sequences of conserved regions of reverse transcriptase (A), protease (B), and integrase (C) domains encoded by Ulysses and other retroviral elements (number of initial codon shown in alignment immediately follows element name). Amino acid sequences (single-letter code) of conserved motifs are in the amino- to carboxyl-terminus direction. Positions at which five or more sequences share identical or chemically similar amino acids are shaded. Gaps were introduced in A to maximize alignment. Chemically similar amino acids are grouped as follows (17): A, S, T, P, and G; N, D, E, and Q; H, R, and K; M, L, I, and V; and F, Y, and W. MoMLV, Moloney murine leukemia virus; aa, any residue.

ment of the Ulysses retrotransposon under controlled conditions during *D. virilis* dysgenesis, either as an effector or as a consequence of this syndrome, will afford the opportunity to study molecular mechanisms underlying the mobilization of LTR-containing retrotransposons in *Drosophila*.

We thank Dr. M. M. Green for critical reading of the manuscript and encouragement and support during these studies. We also thank Dr. L. Rabinow for providing *D. melanogaster* white gene probes. This work was supported, in part, by U.S. Public Health Service Award GM 35463 and American Cancer Society Grant NP-546A to V.G.C.

- Kidwell, M. G., Kidwell, J. F. & Sved, J. A. (1977) *Genetics* **86**, 813–833.
- Hiraizumi, Y. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 268–270.
- Hiraizumi, Y. (1977) *Genetics* **87**, 83–89.
- Rubin, G. M., Kidwell, M. G. & Bingham, P. M. (1982) *Cell* **29**, 987–994.
- O'Hare, K. & Rubin, G. M. (1983) *Cell* **34**, 25–35.
- Finnegan, D. J. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 503–518.
- Rio, D. C., Laski, F. A. & Rubin, G. M. (1986) *Cell* **44**, 21–32.
- Mizra, S. & Rio, D. C. (1990) *Cell* **62**, 269–284.
- Blackman, R. K., Grimaila, R., Koehler, M. D. & Gelbart, W. M. (1987) *Cell* **49**, 497–505.
- Lozovskaya, E. R. & Evgen'ev, M. B. (1987) *Dokl. Akad. Nauk SSSR* **296**, 727–731.
- Lozovskaya, E. R., Scheinker, V. S. & Evgen'ev, M. B. (1990) *Genetics*, in press.
- Zelentsova, E. S., Vashakidze, R. P., Kraev, A. S. & Evgen'ey, M. B. (1986) *Chromosoma* **93**, 469–476.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Bonner, J. J. & Pardue, M. L. (1976) *Chromosoma* **58**, 87–99.
- Gubenko, I. S. & Evgen'ev, M. B. (1984) *Genetica* **65**, 127–139.
- Schwartz, R. M. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, pp. 353–358.
- Yuki, I., Inouye, S., Ishimaru, S. & Saigo, K. (1986) *Eur. J. Biochem.* **158**, 403–410.
- Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. & Yuki, S. (1984) *Nature (London)* **312**, 659–661.
- Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5**, 1630–1638.
- Marlor, R. L., Parkhurst, S. M. & Corces, V. G. (1986) *Mol. Cell. Biol.* **6**, 1129–1134.
- Inouye, S., Yuki, S. & Saigo, K. (1986) *Eur. J. Biochem.* **154**, 417–425.
- Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
- Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) *Nature (London)* **326**, 662–669.
- Bingham, P. M. & Zachar, Z. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 485–502.
- Engels, W. R. & Preston, C. R. (1984) *Genetics* **107**, 657–678.
- Yannopoulos, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P. & Louis, C. (1987) *Cell* **49**, 487–495.
- Lim, J. K. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9153–9157.
- Patterson, J. T. & Stone, W. S. (1952) *Evolution in the Genus Drosophila* (Macmillan, New York).
- Eggleston, W. B., Johnson-Schlitz, D. M. & Engels, W. R. (1988) *Nature (London)* **331**, 368–370.
- Gerasimova, T. I., Matyunina, L. V., Ilyin, Y. V. & Georgiev, G. P. (1984) *Mol. Gen. Genet.* **194**, 517–522.
- Mével-Ninio, M., Mariol, M. C. & Gans, M. (1989) *EMBO J.* **8**, 1549–1558.
- Vashakidze, R., Zelentsova, E., Korochkin, L. I. & Evgen'ev, M. B. (1989) *Chromosoma* **97**, 374–380.
- Geyer, P. K., Richardson, K. L., Corces, V. G. & Green, M. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6455–6459.
- Medhora, M. M., Macpeek, A. H. & Hartl, D. L. (1989) *EMBO J.* **7**, 2185–2191.