Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 719–722, January 1994
Genetics

Nature screen: An efficient method for screening natural populations of *Drosophila* for targeted *P*-element insertions

(transposon tagging/reverse genetics/PCR/mutagenesis)

ANDREW G. CLARK*, SCOTT SILVERIA†, WENDY MEYERS†, AND CHARLES H. LANGLEY†

*Department of Biology, Pennsylvania State University, University Park, PA 16802; and †Center for Population Biology, University of California, Davis, CA 95616

Communicated by M. M. Green, October 1, 1993

ABSTRACT **The efficiency of molecular techniques is making it increasingly necessary to rely on reverse genetics to understand the function of genes. Tissue-specific libraries allow one to identify numerous genes that can be cloned, sequenced, and mapped and whose temporal and tissue-specific pattern of expression are well characterized but whose function remains unknown. In such cases, it is desirable to generate targeted mutations to examine the phenotype of loss-of-function lesions. Here we describe a method for identifying naturally occurring variants of *Drosophila melanogaster* with specific genes tagged by a nearby *P* element. Imprecise *P*-element excision can then be used to generate a series of small deletions in or near the gene. In the method described here, large numbers of wild-caught males were crossed to balancer females, and inserts were identified in pooled samples by the polymerase chain reaction with one primer from each target gene and one primer from the *P*-element terminal repeat. We present the calculations for the probability of successfully tagging a gene and show that it is greatly improved by simultaneously screening inserts into several genes. If a large natural population is available, a nature screen is faster and easier than inducing *P*-element transposition in the laboratory, but the resulting lines, being genetically heterogeneous, may require more subsequent work to isolate. Using this method to screen the genomes of ≈10,400 males, we found *P*-element inserts in close proximity to 3 of 10 genes that were screened.**

It is frequently desirable to generate mutations that alter or eliminate the expression of genes for which the only information available is the nucleotide sequence. This situation arises when genes are cloned from tissue-specific libraries or when they are identified by homology with some other gene. Two methods have been published that tag targeted genes in *Drosophila melanogaster* by mobilizing transposable *P*-elements and by employing the polymerase chain reaction (PCR) to identify inserts near the gene (1, 2). The first makes use of the fact that PCR will amplify a particular sequence even if it is present in <1% of the genomes from which the template DNA is prepared. The chance that any single *P*-element insertion will transpose into the neighborhood of a targeted gene is roughly the size of the targeted region divided by the genome size, a figure on the order of 1/100,000. To avoid having to perform astronomical numbers of PCRs, Ballinger and Benzer (1) pooled the genomic DNA isolated from 100 males with independent *P*-element insertions. PCR with one primer in the targeted gene and one primer on the *P*-element terminal repeat produces a fragment only if the *P* element had inserted near the gene (Fig. 1). If a band is seen, it implies that one of the 200 haploid genomes within the DNA sample has the appropriate insert. It is then a matter of sequentially subdividing the descendants of the original 100 males that founded each pooled population until a pure line is isolated. An alternative approach is to perform the PCR screening on embryos, retaining females in sequentially smaller batches until the desired line is identified (3).

FIG. 1. Diagram of the method by which *P* elements near a target gene are detected. Single oligonucleotides prime DNA synthesis within each target gene and from the *P*-element terminal repeats. T1L and T1R are the left and right primers for target gene 1. Target gene primers are oriented toward each other, so that *P*-element insertion within the gene will be detected. Only if a *P* element is within about 2 kb of a target gene will geometric amplification of a DNA fragment result, as indicated by the heavy line.

As simple and elegant as this method is, it has had only modest success, largely because of the high rate of false positives from the PCR. This prompted Hamilton *et al.* (2) to suggest an alternative approach in which the *P* elements that are mobilized have a bacterial plasmid origin of replication and a carbenicillin-resistance marker. After mobilizing such elements in many independent crosses, the elements are stabilized by genetic removal of the Δ2–3 transposase source. Genomic DNA is prepared from flies pooled across lines, partially digested with *Eco*RI, and ligated under conditions that favor circularization. Competent *Escherichia coli* cells are then transformed with this DNA and selected on carbenicillin. The resulting plasmid-rescue library is then labeled by nick-translation and used as a hybridization probe to screen for sequence similarity with a battery of cDNA clones that have been bound to nylon membranes. While this method seems more technically difficult than that of Ballinger and Benzer (1), the plasmid rescue approach is very efficient for screening certain large classes of inserts (e.g., brain cDNAs). One disadvantage, relevant to the method we report, is the requirement for the use of engineered *P* elements bearing appropriate markers and plasmid functions. The method of Hamilton *et al.* (2) is very labor intensive and appears best suited for the situation in which one does not require an insertion in a particular gene, but where insertions in any of a large class of genes are informative and useful.

The key difference between our method and that of Ballinger and Benzer (1) is that our approach obviates the need for doing the crosses to mobilize *P* elements in a laboratory stock. Instead, we make use of the fact that transposable elements are found in natural populations nearly uniformly distributed throughout the genome (4). Even though there may be 30–50 copies of the *P* element per genome of wild-caught flies, the site occupancy is low, meaning that multiple recovery of the same insertion is unlikely. Naturally occurring *P* elements represent a diverse collection of defective or partially deleted elements whose insertional prop-

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.

erties may vary, but defective elements generally retain the terminal inverted repeats, so they can still be used to prime the PCRs and identify tagged sites. Because most incomplete elements have intact terminal repeats, they can be mobilized to generate imprecise excisions once the insertion is identified. There has been a suggestion that Δ2-3-induced *P*-element transposition produces different arrays of insertions depending on the initial line used, and this clustering of insertions does not appear to be present among naturally occurring insertions (4). Hence, nature has already done the *P*-element mobilization, and all that is needed is to collect the flies, allow them to reproduce, and test them for inserts.

There is considerable gain in efficiency that arises when the PCRs are multiplexed with multiple primers in each reaction, and we devised a scheme for identifying sets of primers to minimize false priming. False priming is readily detected by repeating PCRs with individual target gene primers and by Southern blot analysis. As in Ballinger and Benzer (1), we apply the PCR to genomic DNA isolated from pooled groups of 100 males that had already reproduced. Oligonucleotide primers complementary to one strand of several target genes were included with an oligonucleotide primer complementary to the terminal repeats of the *P* element. If a *P* element had inserted within about 2 kb of the target gene, the pair of primers produced a PCR product that was identified on an agarose gel. We get around the serious problem of false positives by performing a Southern blot analysis of the suspected PCR products. There is considerable advantage accrued to screening for inserts in or near several genes at the same time, and this fact is best appreciated by some simple calculations.

With a few simple assumptions, one can estimate the probability that such a procedure ought to work. Assume the PCR will generate a product if the *P* element is within a region of $t = 2$ kb from the target primer site. With a haploid genome size of $g \approx 170{,}000$ kb (5) and $c \approx 20$ *P* elements per genome, the probability of a successful hit is $s = tc/g$ or $\approx 10^{-4}$ per haploid genome. The chance that there will be no inserts after n diploid flies are screened for a single gene is $(1 - s)^{2n}$. If the PCR is done on i distinct genes, the probability of no hits near any of the target genes is $(1 - s)^{2in}$. Fig. 2 shows that by

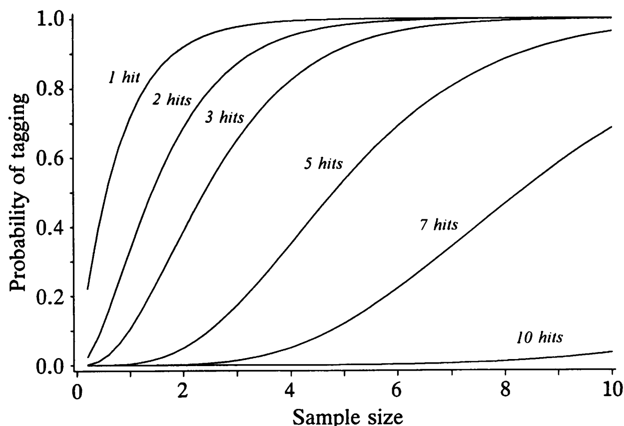

FIG. 2. Probability of successfully identifying at least one *P* element near a target gene depends on the genome size (g), the size of the region that PCR will amplify reliably (t), the number of *P* elements inserted per genome (c), the sample size of diploid genomes surveyed (n), and the number of potential target genes (i) according to the formula $P = 1 - (1 - tc/g)^{2in}$. The figure represents the case of $g = 170{,}000$ kb; $t = 2$ kb; $c = 20$ inserts per genome, in a screen for $i = 10$ genes and shows the probability of at least one hit, at least two hits, etc., as a function of sample size. By performing PCRs on pooled samples, the number of primary PCRs can be as low as 100 and the probability of success in tagging will still be very high. Sample size is shown as no. $\times 10^{-3}$.

looking for *P* elements near any of several genes, the chance that at least one is tagged is very high with just a few thousand genomes surveyed. Given the approach of pooling DNA from 100 males (200 haploid genomes) per initial PCR, excellent prospects for success should be secured with only 100 primary PCRs.

MATERIALS AND METHODS

Fly Collection and Genetic Crosses. During the period from mid-July to mid-August 1992, *Drosophila* were collected in about 50 plastic buckets partly filled with mashed rotting peaches and left under the peach trees in the Wolfskill orchard of the University of California near Winters, California. Flies were collected in the early morning hours and again just before sunset. In the laboratory, male *D. melanogaster* were separated from the females and from *Drosophila simulans* (which were about 50% the sample at this time of year) and were placed in groups of 20 into half-pint bottles with 20–30 virgin *T*(2;3) *CyO*; *TM2*/+; *ry*, *Sb* females. Note that the F_1 progeny will have *P* elements mobilized (6), but the rate of excision was low enough that this did not appear to be a problem. The virgin females were between 1 and 5 days old and mated readily with the wild-caught males. It is important that there be little opportunity to express differential mating success by the males, because we want an even representation of all 20 males in the progeny within a bottle.

Genomic DNA Extraction. After 2–4 days, males were removed and genomic DNA was prepared from groups of 100 males collected from five bottles. Flies were homogenized in a Dounce homogenizer in 2 ml of nuclear isolation buffer (NIB = 0.15 M NaCl/0.01 M Tris·HCl, pH 8.0/0.005 M EDTA/0.2% Nonidet P-40). The homogenate was filtered through a large pipet tip with glass wool and centrifuged at 7000 rpm in a SS34 rotor for 5 min. Supernatant was aspirated off, and pelleted nuclei were resuspended in 400 μl of NIB buffer. Nuclei were then lysed in 1.2 ml of lysis buffer (0.3 M NaCl/0.05 M Tris·HCl, pH 8.0/0.005 M EDTA/1% sodium sarkosyl) and gently mixed. Two rounds of phenol/chloroform extraction were performed, and DNA was precipitated with ethanol and desiccated before being resuspended in TE (50 mM Tris·HCl, pH 8.0/20 mM EDTA).

Design of Multiplexed PCR. The key to the efficient recovery of targeted *P*-element inserts is that a PCR can amplify the appropriate fragment even when multiple primers are simultaneously used in a reaction mixture in which the target sequence is present in only 1 genome of the 200 genomes in a DNA preparation. There were 10 genes that our laboratories were interested in tagging, and since *P*-element inserts on either side of each gene were of interest, 20 target gene primers were used. Primers were designed from the published sequences of *Acp26Aa* (7), *Acp95E* (8), and *Mlc1* (9) and from sequences of genes for accessory gland proteins *Acp29B, Acp33A, Acp36DE, Acp53E, Acp63F, Acp76A, and Acp98B* kindly provided by Mariana Wolfner (Cornell University). We used the following strategy for identifying combinations of primers that would work well together. It is important to plan these combinations carefully, because as more genes are screened, the chance that some combination of primers will dimerize or cause other artifacts increases. We used Bill Engels' (University of Wisconsin) program called AMPLIFY to identify all pairs of primers that present a risk of dimerizing at the PCR annealing temperature. AMPLIFY also allowed us to find all cases in which a primer may weakly hybridize to the wrong gene. We wrote a program to produce lists of primer combinations that satisfy the following three criteria: (*i*) no primer dimers, (*ii*) no false priming, and (*iii*) no pairs of primers for the same gene in the same PCR mixtures. We used five groups of four target primers, and, while five or six primers may work in the same reaction, we got the most

consistent results by limiting the number of target gene primers to four. We used the primer sets that satisfied these criteria on genomic fly DNA and chose a set that had no artifact bands. Our positive control was to use genomic DNA from a white^{hd} stock (6) and an appropriate white primer to give a fragment in combination with the *P*-element primer.

PCR Amplification and Line Identification. Each genomic DNA sample was amplified in five PCRs, each of which included the *P*-element terminal repeat primer and four additional primers. The *P*-element terminal repeat primer (5′-GCGGCCGCGACGGGACCACCTTATGTT-3′) had an 8-bp *Not* I site at its 5′ end that was not used in this study. This oligonucleotide primes DNA synthesis in the direction away from the *P* element. PCR amplification was performed in a total reaction mixture of 25 μl consisting of 2 μg of genomic DNA, 75 ng of the *P*-element primer, each of four primers (each at 50 ng), 1 unit of *Taq* polymerase, all four dNTPs (each of 0.2 mM) in PCR buffer (60 mM Tris·HCl, pH 8.5/2 mM $MgCl_2$/0.02% gelatin/0.0007% 2-mercaptoethanol/0.2% Triton X-100). The amplifications were done with a 1-min denaturation at 94°C, followed by 30 cycles with a 45-sec denaturation at 92°C, a 45-sec hybridization at 60°C, and a 3-min elongation at 72°C. If an amplification product was detected, the PCR was repeated with individual target primers (paired with the *P*-element terminal repeat primer), and these PCR products were separated on an agarose gel and examined by Southern blot analysis. If a product was unambiguously seen for one of the primers, the relevant line was isolated as described below.

Southern Blot Analysis. Southern blot analysis was performed to verify that the bands seen in the PCRs were actually homologous to the appropriate target gene. Southern blot analysis was performed only if there was a putative positive on the primary PCRs. Standard protocols were used for Southern blot analysis (10) and probes were generated by PCR amplification of the targeted genes. High stringency (57°C) was used in washing the blots, because the probes should precisely match the blotted PCR products. True positives gave a very strong signal (Fig. 3).

Isolation of Pure Tagged Lines. If a sample was found to definitively give a PCR product that bound the appropriate probe in a Southern blot, the problem remained to isolate the insertion line from the heterogeneous set that composed the original sample of genomic DNA. First, genomic DNAs were extracted from about 100 F_1 larvae and/or emerging F_1 adults from each of the five bottles from which the male parents were pooled. PCR then identified which one of the five bottles contained the positive line. Within the identified bottle, there may be as many as 40 haploid lineages, so it was necessary to establish a large set of lines to recover the desired insert. Males from the bottle were crossed to virgin *T(2;3) CyO*; *TM2/+*; *ry, Sb* females as single lines in about 100 vials. These were then tested individually by PCR. In case of any ambiguities in the PCR gels, identities of bands were verified by Southern blot analysis. After one or two rounds of this process, pure lines were obtained.

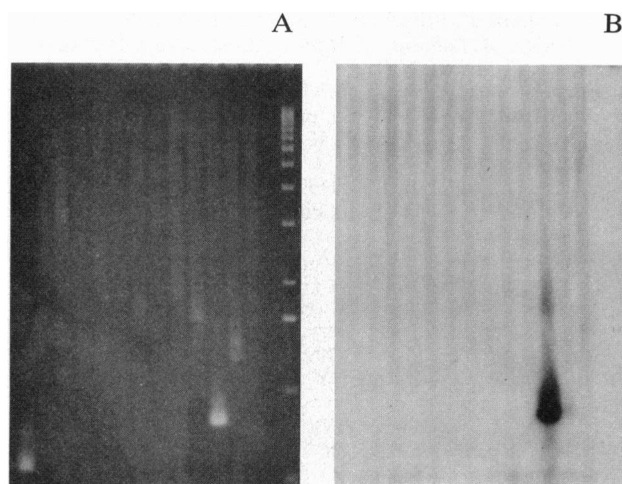

FIG. 3. (*A*) Ethidium bromide-stained agarose gel showing the results of PCRs using the *P*-element primer and a primer from the accessory gland protein *Acp36DE* on genomic DNA extracted from individual bottles containing 40 haploid genomes. Note the occurrence of many faint bands (false positives). The rightmost lane is a size standard. (*B*) Southern blot of the same gel, probed with the PCR product formed with the two *Acp36DE* primers. The bright band on the agarose gel in *A* corresponds to the pronounced signal on the Southern blot.

RESULTS

In July and August 1992, 10,400 males were caught in the field, 520 bottles were set up, and 104 genomic DNA samples were extracted from pooled sets of five bottles. Five PCRs were done on each of these DNA samples and resulting products were visualized after agarose electrophoresis. A total of 71 putative positives were identified from these primary amplifications. Genomic DNA was extracted from the F_1 progeny in each bottle of the appropriate sets (after they had reproduced), and in 23 cases, the PCRs of the progeny continued to reveal putative inserts. Southern blot analysis confirmed that these were true positives in five cases (Fig. 3), and in each case, F_1 progeny from each bottle were distributed into 100 vials, male F_1 progeny were crossed again to *T(2;3) CyO*; *TM2/+*; *ry, Sb* females. One of these gave bands from heterogeneous cultures, but a homozygous line could not be produced, apparently due to sterility. In another case, a PCR fragment appeared sporadically in a lineage, but we were unsuccessful in establishing a pure line. In this case the insert was in low frequency in the bottle, and by either drift or selection, it was lost before a vial culture could be established. Eventually three *P*-element insertion lines were established with inserts near 3 of the 10 initially targeted genes. The genes that were tagged include the three accessory gland genes, *Acp36DE, Acp95E*, and *Acp98B*.

DISCUSSION

Potential Difficulties with Nature Screen. Despite our success in isolating three lines of flies with selected loci tagged by *P* elements, there are aspects of the nature screen that may limit its utility. As in the method of Ballinger and Benzer (1), once one has a line with the desired *P*-element insert, further work is required to remove other *P* elements from the genome. If this is not done, crosses designed to excise the targeted *P*-element insert will generate many other transposition and rearrangement events. The genetic background of the flies will be highly heterogeneous. If one wishes to test subtle phenotypic effects of insertional nulls, more work will be necessary to remove this background heterogeneity. Sometimes the presence of lethals and steriles in the genetic background may make it very difficult to isolate a homozygous line with an insert of interest. The method also requires primers for each targeted gene, and for some problems this may either be impossible or prohibitively expensive.

Other problems may arise because of the logistics of carrying out a large-scale experiment that depends on a natural population. If the population has a short period of peak numbers, the experiment must be done within the window of opportunity. It is necessary to have the results of the primary PCRs before the progeny of the first generation cross have expired, and this means that one must be prepared for a fairly intensive session of PCRs. We had three PCR

machines running for much of a 2-week period. If the fly season is of longer duration, then a more efficient strategy might be to set up cultures directly with wild-caught inseminated females. After the females have laid eggs, their DNA can be extracted in groups and assayed by PCR. This sequential approach was used in a trial run, and it seemed applicable to a less-intensive and more-drawn-out screen.

Advantages of Nature Screen. The method described here differs from previous approaches to gene-specific *P*-element mutagenesis in two ways. (*i*) Difficult time-consuming *Drosophila* crosses to generate mobilized *P* elements were replaced by healthful fun field collecting. The cross involving mobilization of many *P* elements simultaneously by the Δ2–3 element, as used in previous tagging methods (1, 2), results in flies of low fertility, compounding the difficulties of generating large numbers of lines. (*ii*) A considerable gain in efficiency is afforded by multiplexing the PCRs to identify inserts in the proximity of multiple genes. Most of the labor in a screen of this type is in collecting and crossing the flies and in preparing genomic DNAs, and one can double the number of genes that are screened with a relatively small increase in effort. The labor involved in screening thousands of genomes by a nature screen is much less than that for the method of Hamilton *et al.* (2), but the two approaches are aimed at different problems. The nature screen is ideally suited to finding an insert near particular genes, whereas the method of Hamilton *et al.* (2) appears to be best suited for the situation in which one wishes to screen many potential genes in a smaller number of genomes, and one seeks clones in a broad class, such as genes expressed in a particular tissue.

The population genetics of transposable elements suggests that our success in tagging three genes was not a fluke. The calculation of the probability of success assumes that *P*-element inserts occur uniformly throughout the genome. Although there is evidence that there are local inhomogeneities in *P*-element insertions (11), the data also clearly demonstrate that the number of occupiable sites is enormous. Every survey of restriction site variation in and around cloned genes in *D. melanogaster* has identified transposable elements in the proximity of the gene. All that is necessary for our scheme to work is that there be many occupiable sites and that the frequency of insertions at each site be low. It is not necessary to know anything about the frequency of insertion and deletion or the action of other evolutionary forces on the elements. Primers designed from other transposable elements could also be used to search for naturally occurring insertions near genes of interest (12). In this study, we chose the *P* element simply because more is known about regulating its transposition for the purposes of generating excisions. But as we learn more about the regulation of transposition of *hobo* (13, 14) and other elements, they too will be amenable to screening for naturally tagged genes.

The philosophy of the nature screen is concordant with a rich history of successes in screening natural populations for important mutations of *Drosophila*. The elucidation of fundamental processes of genetics by Morgan, Bridges, Sturtevant, and others depended on identification of naturally occurring and spontaneous mutants. Our understanding of gene expression has been enriched by studies of diverse alleles of white (15) and other loci (16) that were obtained by screening natural populations. Null alleles of several enzyme-encoding loci were recovered with a large electrophoretic screen (17). The discoveries of Segregation Distorter (18) and many other meiotic mutants (19) were made through large-scale screens of natural populations. The original characterization and ultimate identification of the *P* element began with surveys of natural populations by several workers. With this track record, it seems prudent to be open to the possibility that natural populations continue to harbor genetic variation essential to understanding contemporary problems in genetics as well.

Note Added in Proof. Michael Bertram and Mariana Wolfner (personal communication) found that the *Acp36DE* insert, which occurred within the gene, causes a truncation of the protein product. Western blot screening demonstrated that the *P* element is mobilizable and that precise and imprecise excisions have been generated.

We thank Mariana Wolfner, Heidi Harada, Michael Bertram, and Kevin Kraus for providing unpublished sequence data; Montserrat Aguadé and Jeanette Robertson for assisting in fly handling; and Phyllis Wescott for preparing medium. Jeanette Natzle and Ken Burtis also gave helpful suggestions as they took part in the screen. This work was supported by National Science Foundation Grant BSR-9117222 to C.H.L. and by Public Health Service Grant GM45876 and a Sabbatical Fellowship from the Alfred P. Sloan Foundation to A.G.C.

1. Ballinger, D. G. & Benzer, S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9402–9406.
2. Hamilton, B. A., Palazzolo, M. J., Chang, J. H., VijayRaghavan, K., Mayeda, C. A., Whitney, M. A. & Meyerowitz, E. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2731–2735.
3. Kaiser, K. & Goodwin, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1686–1690.
4. Charlesworth, B. & Langley, C. H. (1989) *Annu. Rev. Genet.* **23**, 251–287.
5. Mulligan, C. D. & Rasch, E. M. (1980) *Histochemistry* **66**, 11–18.
6. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, A. J. (1990) *Cell* **62**, 515–525.
7. Monsma, S. A. & Wolfner, M. F. (1988) *Genes Dev.* **2**, 1063–1073.
8. DiBenedetto, A. J., Harada, H. A. & Wolfner, M. F. (1990) *Dev. Biol.* **139**, 134–148.
9. Falkenthal, S., Parker, V. P. & Davidson, N. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 449–453.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
11. Engels, W. R. (1989) in *Mobile DNA*, eds. Berg, D. & Howe, M. (Academic, New York), pp. 437–484.
12. Milligan, C. D. & Kaiser, K. (1993) *Nucleic Acids Res.* **21**, 1323–1324.
13. Pascal, L. & Periquet, G. (1991) *Mol. Biol. Evol.* **8**, 282–296.
14. Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. (1991) *Cell* **66**, 465–471.
15. Green, M. M. (1959) *Proc. Natl. Acad. Sci. USA* **45**, 549–553.
16. Spencer, W. P. (1947) *Adv. Genet.* **1**, 359–402.
17. Voelker, R. A., Langley, C. H., Leigh Brown, A. J., Ohnishi, S., Dickson, B., Montgomery, E. & Smith, S. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1091–1095.
18. Sandler, L. M., Hiraizumi, Y. & Sandler, I. (1959) *Genetics* **44**, 233–250.
19. Sandler, L. M., Lindsley, D. L., Nicoletti, B. & Trippa, G. (1968) *Genetics* **60**, 525–558.