# The Transposable Portion of the Genome of *Drosophila algonquin* Is Very Different from That in *D. melanogaster*[1]

*Jody Hey*

Department of Ecology and Evolution, State University of New York at Stony Brook

Four clones containing different transposable elements were isolated from a genomic library of *Drosophila algonquin*. Each clone was hybridized to salivary-gland chromosomes of three lines of *D. algonquin* and two lines of *D. affinis*. The estimated copy number in *D. algonquin* of the four element families varied from 59 to 333. The occupancy per site varied from 0.64 to 0.75. Thus the transposable portion of the *D. algonquin* genome is dominated by a few high-copy-number elements, each characterized by high occupancies. The copy number and occupancy values were very similar in *D. affinis*. This differs from the situation in *D. melanogaster* mobile middle-repetitive DNA, which has at least 30 and perhaps as many as 100 different families of mobile elements, with copy numbers ranging from 5 to 100. When several lines have been examined, elements in *D. melanogaster* are revealed to have very low occupancies. The four *D. algonquin* elements do not hybridize with *D. melanogaster* DNA, but they did hybridize with 15 *obscura*-group species, thereby revealing a pattern that is consistent with concerted evolution.

## Introduction

The presence of families of mobile genetic elements in the genomes of many organisms raises such questions as the following: How many families are there? How many elements are there in each family? and, To what degree do elements occupy the same positions in the DNA of every individual?

*Drosophila melanogaster* is thought to carry between 30 and 100 families, with a range of 10–100 copies/family (Manning et al. 1975; Young 1979; Dowsett and Young 1982; Finnegan and Fawcett 1985). The locations of elements vary widely among different lines of *D. melanogaster* (Strobel et al. 1979; Montgomery and Langley 1983; Ananiev et al. 1984; Belyaeva et al. 1984; Ronsseray and Anxolabehere 1986; Leigh Brown and Moss 1987).

I investigated transposable elements in *D. algonquin* and found marked differences from *D. melanogaster* in both copy number and site occupancy.

## Material and Methods

### *Drosophila* Lines Used for In Situ Hybridization

Each line was derived from a single female collected between 1983 and 1984. The lines had undergone ≥20 generations in small cultures followed by ≥2 generations of full-sib inbreeding and thus were expected to be inbred. The *D. algonquin* lines were ALli from Long Island; ALbi from Bull Island, N.J.; and ALmc from Machias,

1. Key words: *Drosophila*, transposable elements, *affinis* subgroup, middle repetitive DNA.

Address for correspondence and reprints: Dr. Jody Hey, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138.

*Mol. Biol. Evol.* 6(1):66–79. 1989.

0737-4038/89/0601-0005$02.00

Maine. The lines of *D. affinis* were AFho from Houston and AFli from Long Island. Flies were reared according to the method of Hey and Houle (1985).

## Library Construction

*Drosophila algonquin* DNA was prepared from adults of line ALli (Bingham et al. 1981), and a 10-μg sample was partially digested with *Mbo*I to an average size of 20 kb. This DNA was electrophoresed in a gel of 0.5% agarose at 1.0 V/cm overnight, and fragments in the size range of 10–20 kb were isolated by electroelution using an IBI Technologies electroeluter. The resulting DNA (~1 μg) was resuspended in 10 μl of $^1/_{10}$ TE 8 [1 mM tris hydroxymethyl aminomethane (Tris)-HCl, pH 8, and 0.1 mM ethylenediaminetetraacetate (EDTA)]. DNA from λ phage vector EMBL-4 (Frischauf et al. 1983) was prepared according to the method of Maniatis et al. (1982, pp. 77–83) and cut with *Bam*HI and *Sal*I. Five-microgram samples of this DNA were dephosphorylated, precipitated in ethanol, and resuspended in 5 μl of $^1/_{10}$ TE 8. The genomic and λ phage DNA solutions were mixed and joined with 1 unit of T4 DNA ligase. One-fifth of the ligation reaction was combined with packaging extracts (Maniatis et al. 1982, pp. 264–268), yielding an estimated 12,000 plaque-forming units.

## Clone Screening

A screening procedure was employed to identify clones carrying repetitive sequences (J. Ajioka and P. Bingham, personal communication). Nitrocellulose was used to lift plaques from each of two square grid plates. Genomic DNA was melted and reannealed to a Cot of 2.0 as follows: 10 μg of *D. algonquin* DNA in 100 μl of TE buffer (10 mM Tris-HCl, pH 8, and 1 mM EDTA) were heated at 90 C for 1 min; 5 μl of 4 M NaCl were added, and the solution was heated at 65 C for 2 h. To remove single-stranded DNA, 100 μl of 2 × S1 nuclease buffer (0.5 M NaCl, 0.1 M Na-acetate pH 4.6, 9.0 mM $ZnSO_4$, and 40 μg salmon sperm DNA/ml) were added, followed by 50 units of S1 nuclease. The mixture was incubated at 37 C for 30 min, following which the DNA was extracted once in phenol/$CHCl_3$ and precipitated with ethanol. On the basis of results from work on *D. melanogaster* (Manning et al. 1975), this DNA was expected to contain most of the middle and highly repeated sequences in the genome. Of the resulting 1.5 μg of DNA, 0.5 μg were nick-translated with $^{32}$P-α-dCTP and were hybridized to the nitrocellulose filters. After autoradiography, plaques corresponding to spots of hybridization on the autoradiograph were removed from a replicate set of plates. The spots of hybridization were ranked in order of intensity, and 15 clones were selected at uniform intervals from this ranking. The plate growth procedure of Helms et al. (1985) was used to prepare DNA from individual clones. DNA from these clones was hybridized in situ to salivary-gland chromosome squashes according to the method of Montgomery et al. (1987).

## Determining the Presence of Cloned Sequences in Other Species

DNA was prepared by the method of McGinnis et al. (1983) from each of the following lines and species: ALli; AFho; *D. melanogaster* FM6; *D. narragansett,* collected by the author; *D. azteca,* collected by D. Houle; *D. athabasca* semispecies EA and EB, collected by the author; *D. athabasca* semispecies WN, provided by A. Beckenbach; *D. pseudoobscura,* National Drosophila Species Resource Center (NDSRC) number 14011-0121.11; *D. pseudoobscura bogotana,* NDSRC number 14011-0121.68; *D. ambigua,* NDSRC number 14011-0091; *D. miranda,* NDSRC number 14011-0101; *D. persimilis,* NDSRC number 14011-0111.1; *D. bifasciata,* NDSRC number

14012-0181; *D. tolteca*,, NDSRC number 14012-0201; and *D. subobscura*, NDSRC number 14011-0131. DNAs were digested with *Pst*I or *Bam*HI and electrophoresed in 0.8% agarose gels at 0.6 V/cm for 15 h. Gels were blotted with nylon membrane, New England Nuclear Gene Screen, and clones that had previously been identified as carrying transposable sequences were nick-translated and hybridized to the membranes.

### Estimating Transposition and Deletion Rates

Two models have been developed to estimate rates of transposition and deletion from data on the occupied sites in a sample of chromosomes. They assume that (1) all sites are effectively unlinked, (2) all insertions are equally subject to deletion, and (3) the probability of replicative transposition from an insertion is a decreasing function of the number of insertions in the genome, thus leading to self-regulation of copy number. Charlesworth and Charlesworth (1983) defined the parameters $\alpha = 4N\mu$ and $\beta = 4N\nu$, where $N$ = effective population size, $\mu$ = the probability of appearance of an additional element copy per generation (by replicative transposition), and $\nu$ = the probability of loss of an individual element copy per generation (by deletion or excision). This model simultaneously provides an estimate of the number of occupiable sites, $T$. Langley et al. (1983) also incorporate the parameter $4N\nu$, $\theta$ in their terminology. Their model assumes that the number of occupiable sites is infinite and that the transposition parameter, $\alpha$, thus can be taken as zero. Estimates of $4N\nu$ for the model of Langley et al. (1983) as determined by the method of Kaplan and Brookfield [1983, equation (7)] will be referred to as $\beta_1$, and the estimate by the method of Charlesworth and Charlesworth (1983) will be referred to as $\beta_2$. $T$, $\alpha$, and $\beta_2$ are calculated from equation (A 9) of Charlesworth and Charlesworth (1983).

## Results

### Description of Transposable Element–bearing Clones

Screening with the repetitive probe identified 128 (26%) of 490 clones as carrying repetitive sequences. In situ hybridization to lines ALli and ALbi revealed that all 15 clones carry DNA homologous to multiple cytological positions. Five of the clones were not considered further; one hybridized to two invariant locations, another hybridized to six invariant locations, and three hybridized strongly to 5–15 places and very weakly to many others. These latter three almost certainly carried mobile sequences but, because of the difficulty of interpreting the many weak spots of hybridization, were not included in the remainder of the study.

Of the remaining 10 which clearly carried transposable sequences—by the criteria of dispersed multiple locations within strains *and* differences in locations among strains (Young 1979)—six appeared to carry sequences homologous to one of the other four, as indicated by equivalent patterns of in situ hybridization. The remainder of the study was limited to four clones: DA60 represented a set of six equivalent clones; DA70 represented two equivalent clones; and DA21 and DA1 appeared to be unique. The salivary glands of at least two larvae were prepared for every probe and line combination, and no variation appeared within lines, consistent with their being inbred. In addition to the dispersed hybridizations on the polytene chromosome arms, each of these clones hybridized strongly to the chromocenter.

The distal tip of chromosome arm BL and the majority of arm BS were selected for detailed analysis because they carry many puffs and constrictions (fig. 1). This restriction was necessary because most probes hybridized to so many places that the polytene banding pattern was considerably obscured. These two portions of the genome

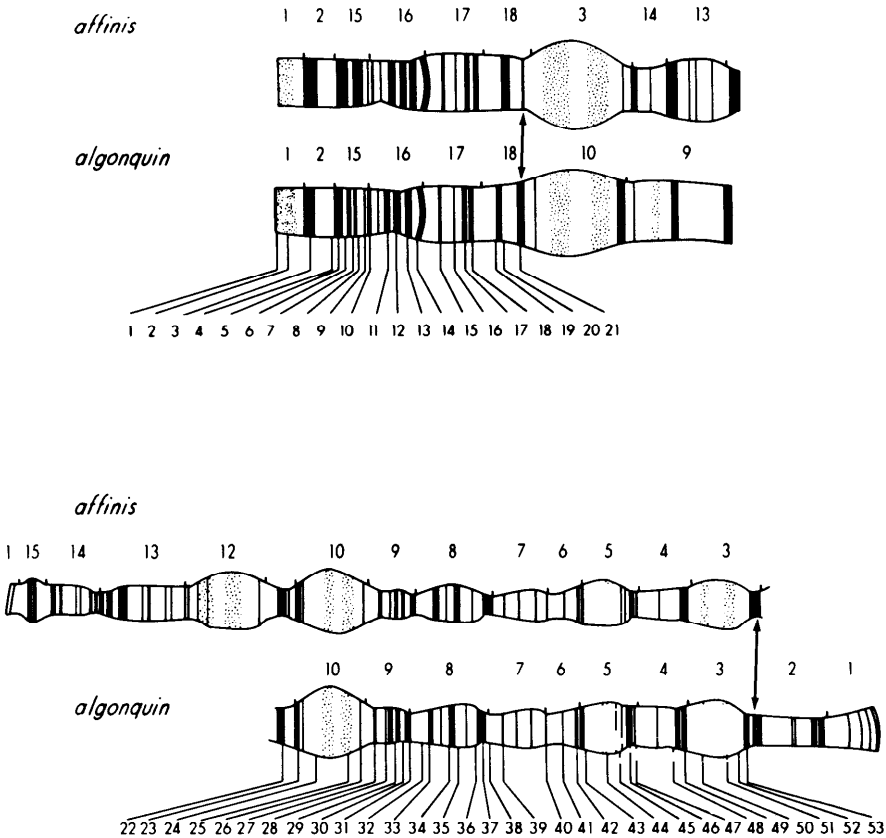


FIG. 1.—Drawings of portions of the polytene chromosomes of *Drosophila algonquin* and *D. affinis*. At the top is shown the distal portion of chromosome BL. At the bottom is shown the distal portion of chromosome BS. The arrows correspond to the endpoints of inversions that differ between the species (Miller 1977). The numbers above the chromosomes correspond to the sections of Miller (1977). The numbers below the chromosomes correspond to the sites of table 1.

make up ~7.3% of the total length of the polytene arms, as measured on the drawings of Miller (1939).

Examination of in situ hybridizations of the four probes to the three *Drosophila algonquin* and two *D. affinis* lines revealed a total of 325 hybridization spots in 53 cytologically distinguishable sites. The location of these sites is shown in figure 1. The data for the location and number of hybridizations of each probe in each of the five lines are shown in table 1. The presence of hybridization by a probe to a site in any of the five lines is called an insertion.

## Analysis of Distribution of Transposable Elements

Table 2 shows the values for several descriptors of insertion distributions. The average number of insertions per line, $\Lambda$, in *D. algonquin* ranges from 4.33 to 24.33. This is a very high density of insertions for only 53 cytologically distinguishable sites. When extrapolated to the full genome, the same values give, for the four elements, estimated total copy numbers that range from 59 to 333 in *D. algonquin* (table 2, col. *G*). The average occupancy per site, $Q$, is >0.6 for all elements in both species.

**Table 1**
**Chromosomal Location of Insertions of Four Transposable Elements**

| Site | Probe and Line | | | | | | | | | | | | | | | | | | | | No. of Probes in Site |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DA1 | | | | | DA21 | | | | | DA60 | | | | | DA70 | | | | | |
| | ALli | ALbi | ALmc | AFho | AFli | ALli | ALbi | ALmc | AFho | AFli | ALli | ALbi | ALmc | AFho | AFli | ALli | ALbi | ALmc | AFho | AFli | |
| A. Chromosome Arm BL | | | | | | | | | | | | | | | | | | | | | |
| 1 .... | — | — | — | * | * | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 1 |
| 2 .... | — | — | — | — | — | * | * | * | — | — | * | * | * | * | * | — | — | — | — | — | 2 |
| 3 .... | — | — | — | — | — | — | — | — | * | * | — | — | — | — | — | — | — | — | * | * | 2 |
| 4 .... | — | — | * | — | — | — | * | * | — | — | — | — | — | — | — | — | — | — | * | * | 3 |
| 5 .... | — | — | — | — | — | — | — | — | — | — | — | * | — | — | — | * | * | * | — | — | 2 |
| 6 .... | — | — | — | — | — | — | — | — | — | — | * | — | — | — | — | — | — | — | — | — | 1 |
| 7 .... | — | — | — | — | — | * | * | * | — | * | * | — | — | * | * | — | — | — | — | — | 2 |
| 8 .... | — | — | — | — | — | — | — | — | * | — | — | — | — | — | * | — | — | — | — | — | 2 |
| 9 .... | — | — | — | — | — | * | * | * | * | — | — | — | — | * | — | * | * | * | — | — | 3 |
| 10 .... | — | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | — | — | — | — | 1 |
| 11 .... | — | — | — | — | — | — | — | — | * | * | — | — | — | — | * | * | * | * | — | * | 3 |
| 12 .... | — | — | — | — | — | — | — | — | — | — | * | — | * | — | — | * | * | * | — | — | 2 |
| 13 .... | — | — | — | — | — | — | — | — | — | — | — | * | * | — | — | — | — | — | — | — | 1 |
| 14 .... | — | — | — | * | — | * | * | * | — | — | * | * | * | — | — | — | — | — | * | * | 4 |
| 15 .... | * | * | * | — | — | — | — | — | — | * | * | * | — | — | * | * | * | * | — | — | 4 |
| 16 .... | — | — | — | — | — | — | * | * | * | * | * | * | * | — | — | * | * | * | * | * | 3 |
| 17 .... | — | — | — | — | — | * | — | — | — | * | — | * | * | * | * | — | — | — | — | * | 3 |
| 18 .... | — | — | — | — | — | — | — | — | — | — | * | — | * | — | — | — | — | — | — | — | 1 |
| 19 .... | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | 1 |
| 20 .... | — | — | — | * | * | * | * | * | * | * | * | * | * | — | — | — | — | * | * | * | 4 |
| 21 .... | — | — | — | — | — | * | * | * | * | — | — | — | — | — | * | — | — | — | — | * | 3 |

B. Chromosome Arm BS

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 .... | — | — | — | — | — | * | * | * | — | * | * | * | * | — | — | * | * | * | — | * | 3 |
| 23 .... | — | — | — | — | — | — | * | * | * | — | * | — | — | — | — | — | * | — | — | — | 3 |
| 24 .... | — | — | — | — | — | — | — | — | — | * | — | — | * | * | — | * | — | * | * | * | 3 |
| 25 .... | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | * | — | — | — | — | 2 |
| 26 .... | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | — | 1 |
| 27 .... | — | — | — | — | — | * | * | * | — | — | * | * | * | — | * | — | * | — | — | * | 3 |
| 28 .... | * | — | * | — | — | * | — | * | * | * | * | * | * | * | * | * | * | * | * | * | 4 |
| 29 .... | — | — | — | — | — | — | — | * | — | — | * | * | * | — | — | — | — | — | — | * | 3 |
| 30 .... | — | — | — | — | — | * | — | — | * | * | * | * | * | — | — | * | * | — | * | * | 3 |
| 31 .... | — | — | — | — | — | * | * | * | — | — | — | * | * | * | * | — | — | * | — | — | 3 |
| 32 .... | — | — | — | — | — | — | — | — | — | — | — | — | — | — | * | * | — | — | — | — | 2 |
| 33 .... | — | — | — | — | — | * | — | * | * | * | * | — | * | — | — | — | — | — | — | — | 2 |
| 34 .... | — | — | — | — | — | * | * | * | — | — | — | * | — | * | — | — | — | — | — | — | 2 |
| 35 .... | — | — | — | — | — | — | — | — | — | — | * | — | — | — | — | — | — | — | — | — | 1 |
| 36 .... | — | — | — | — | — | — | — | — | — | — | * | — | * | — | — | — | — | — | — | — | 1 |
| 37 .... | — | — | — | — | — | * | — | — | — | — | — | * | — | * | * | — | — | — | — | — | 2 |
| 38 .... | — | — | — | — | — | * | — | — | — | * | — | * | — | — | — | — | — | — | — | — | 2 |
| 39 .... | — | — | * | — | — | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | 2 |
| 40 .... | * | * | * | — | — | — | — | * | — | — | — | — | — | — | — | * | * | * | — | — | 3 |
| 41 .... | — | — | — | — | — | * | * | * | * | * | * | * | — | — | — | * | * | * | — | * | 3 |
| 42 .... | — | — | — | — | — | — | — | — | — | — | * | * | * | * | * | — | — | — | — | — | 1 |
| 43 .... | — | — | — | — | — | — | — | — | — | — | — | — | — | * | — | — | — | — | — | — | 1 |
| 44 .... | — | — | — | — | — | — | — | — | * | * | * | * | * | — | * | — | — | — | — | — | 2 |
| 45 .... | — | — | — | — | — | — | — | * | — | — | — | — | * | — | — | — | — | — | * | * | 3 |
| 46 .... | — | — | — | — | — | * | * | * | * | * | — | — | — | * | — | — | — | — | * | — | 3 |
| 47 .... | — | — | — | — | — | — | * | * | — | * | * | — | — | — | — | * | * | * | * | * | 3 |
| 48 .... | — | — | — | — | — | — | — | * | * | * | — | — | * | * | * | — | — | — | — | — | 2 |
| 49 .... | — | — | — | — | — | * | * | * | * | * | * | * | — | — | — | * | * | * | * | * | 3 |
| 50 .... | — | — | — | — | — | — | * | * | — | — | * | * | — | — | * | * | — | — | — | * | 3 |
| 51 .... | * | * | * | — | — | * | — | * | * | * | * | — | * | — | * | * | * | * | * | * | 4 |
| 52 .... | — | — | — | — | — | — | — | * | * | * | — | — | * | * | — | * | * | * | * | * | 3 |
| 53 .... | — | — | — | — | — | * | * | * | * | * | * | * | — | — | * | — | — | — | * | * | 3 |
| Sum | 4 | 3 | 6 | 3 | 2 | 21 | 19 | 27 | 19 | 22 | 26 | 23 | 24 | 15 | 18 | 19 | 18 | 19 | 15 | 22 | 127 |

NOTE.—The presence of hybridization within a site is denoted by an asterisk. The locations of the sites are indicated in fig. 1.

**Table 2**
**Statistics of Transposable Elements in *Drosophila***

| Probe | $n$ | $K$ | $G$ | $\Lambda$ | $Q$ | $\beta_1$ | $\alpha$ | $\beta_2$ | $T$ |
|---|---|---|---|---|---|---|---|---|---|
| DA1: | | | | | | | | | |
| *D. algonquin* | 3 | 6 | 59 | 4.3 | 0.75 | 0.33 | 1.78 | 1.21 | 7.3 |
| *D. affinis* | 2 | 3 | 34 | 2.5 | 0.82 | 0.25 | 3.0 | — | 3.3 |
| DA21: | | | | | | | | | |
| *D. algonquin* | 3 | 31 | 305 | 22.3 | 0.73 | 0.33 | 1.71 | 1.07 | 36.4 |
| *D. affinis* | 2 | 26 | 280 | 20.5 | 0.76 | 0.37 | 1.73 | — | 32.4 |
| DA60: | | | | | | | | | |
| *D. algonquin* | 3 | 38 | 333 | 24.3 | 0.64 | 0.54 | 0.73 | 0.99 | 57.3 |
| *D. affinis* | 2 | 25 | 226 | 16.5 | 0.66 | 1.06 | 0.0 | — | 27.3 |
| DA70: | | | | | | | | | |
| *D. algonquin* | 3 | 26 | 255 | 18.7 | 0.75 | 0.33 | 1.79 | 1.41 | 33.4 |
| *D. affinis* | 2 | 23 | 253 | 18.5 | 0.78 | 0.32 | 2.11 | — | 27.3 |
| *D. melanogaster* (Montgomery and Langley 1983): | | | | | | | | | |
| *copia* | 20 | 27 | 8.0 | 1.6 | 0.06 | 48.3 | 0.23 | 48.4 | 339 |
| 412 | 20 | 40 | 12.5 | 2.5 | 0.06 | 35.0 | 0.00 | 30.1 | infinite |
| 297 | 20 | 51 | 19.0 | 3.8 | 0.07 | 16.7 | 0.05 | 16.7 | 1340 |
| *D. melanogaster* (Leigh Brown and Moss 1987): | | | | | | | | | |
| *copia* | 20 | 31 | 11.3 | 2.3 | 0.07 | 18.3 | 0.80 | 35.0 | 103 |
| *I* | 20 | 46 | 16.0 | 3.2 | 0.07 | 21.5 | 0.87 | 42.5 | 160 |

NOTE.—$n$ = number of inbred lines examined. $K$ = number of sites to which a probe binds in at least one line of a species. $\Lambda$ = mean number of insertions per line within a species and is calculated from equation (2) of Kaplan and Brookfield (1983); $G = \Lambda$ extrapolated to the entire genome, under the assumption that the remainder of the genome has the same density of sites as do the portions included in the study. $G = \Lambda/0.073$ for the *affinis* subgroup species, and $G = \Lambda/0.2$ for *D. melanogaster* (if it is assumed that the X chromosome makes up ~20% of the genome in *D. melanogaster*). $Q$, the mean occupancy per site, is the proportion of lines bearing insertions, averaged among those sites in which at least one insertion was observed in a species. $Q = \Lambda/K$. Transposition and deletion parameters ($\alpha$, $\beta_1$, $\beta_2$, and $T$) are described in Material and Methods. $\beta_2$ cannot be calculated for *D. affinis* because of the small sample size.

In table 3 are presented the observed and expected occupancy profiles for the four probes in *D. algonquin*. Both model A, from Langley et al. (1983) according to the method of Kaplan and Brookfield (1983), and model B, from Charlesworth and Charlesworth (1983), provide close-fitting expectations. Similarly, when these models are applied to data from transposable elements in *D. melanogaster*, they both generate expectations similar to the observed distributions (Charlesworth and Charlesworth 1983; Kaplan and Brookfield 1983; Leigh Brown and Moss 1987). In *algonquin* and *affinis* the estimated number of available sites, $T$, under simultaneous estimation of $\alpha$ and $\beta_2$, is generally close to the actual number of observed sites, $K$. In contrast, estimates of $T$ from *D. melanogaster* are much higher than $K$ (table 2).

Copy number and occupancy estimates are made by assuming that every site of in situ labeling represents a single transposable element, but it is possible that some sites may contain more than one copy of the hybridizing element. One can make a Poisson correction for this under plausible assumptions, though estimates of copy number and site occupancy are not greatly affected.

The pairs of values in the two *affinis*-subgroup species exhibit significant correlations across the four probes for all of the occupancy statistics (for $Q$, $r = 0.965$, $P < 0.05$; for $\beta_1$, $r = 0.993$, $P < 0.01$; and for $\alpha$, $r = 0.923$, $P < 0.1$). The correlation for $\beta_2$ cannot be calculated because the small sample of two lines of *D. affinis* invariably

**Table 3**
**Occupancy per Site in *Drosophila algonquin***

| | No. of Lines with Site Occupied | | | |
|---|---|---|---|---|
| | One | Two | Three | $\chi^2$ |
| DA1: | | | | |
| Observed no. of sites | 2 | 1 | 3 | |
| Expected no. of sites: | | | | |
| A. $\beta_1 = 0.325$ | 1.82 | 1.37 | 2.81 | 0.13 |
| B. $\alpha = 1.78, \beta_2 = 1.21$ | 1.75 | 2.20 | 2.77 | 0.71 |
| C. $\Delta = 0.162, \Delta^{-1} = 6.2$ | 1.15 | 2.71 | 2.14 | 2.05 |
| DA21: | | | | |
| Observed no. of sites | 9 | 8 | 14 | |
| Expected no. of sites: | | | | |
| A. $\beta_1 = 0.328$ | 9.44 | 7.11 | 14.45 | 0.15 |
| B. $\alpha = 1.71, \beta_2 = 1.07$ | 8.23 | 10.76 | 13.30 | 0.82 |
| C. $\Delta = 0.031, \Delta^{-1} = 31.8$ | 5.99 | 14.04 | 10.97 | 4.95 |
| DA60: | | | | |
| Observed no. of sites | 14 | 13 | 11 | |
| Expected no. of sites: | | | | |
| A. $\beta_1 = 0.538$ | 15.47 | 10.06 | 12.47 | 1.17 |
| B. $\alpha = 0.73, \beta_2 = 0.99$ | 14.21 | 12.36 | 11.24 | 0.04 |
| C. $\Delta = 0.025, \Delta^{-1} = 40.7$ | 11.74 | 17.53 | 8.73 | 2.20 |
| DA70: | | | | |
| Observed no. of sites | 10 | 2 | 14 | |
| Expected no. of sites: | | | | |
| A. $\beta_1 = 0.333$ | 7.99 | 6.00 | 12.00 | 3.51 |
| B. $\alpha = 1.79, \beta_2 = 1.41$ | 8.70 | 10.10 | 12.70 | 6.82 |
| C. $\Delta = 0.037, \Delta^{-1} = 26.7$ | 5.10 | 11.80 | 9.11 | 15.47 |

NOTE.—The numbers in the body of the table are the number of sites occupied by an element in one, two, and three of the lines of *D. algonquin* examined. The expected values for model A were calculated from equation (11) of Kaplan and Brookfield (1983). The values for model B were calculated from equation (A 9) of Charlesworth and Charlesworth (1983). The estimates of $\alpha$ and $\beta$ are taken from table 2. For model C, the expected number of sites and the $\Delta$ values are calculated from equations (16) and (17) of Kaplan and Brookfield (1983).

yields $\beta_2 = 1.0$. These correlations must be considered with caution because the different parameters are not independent, and three of the probes form a tight cluster. However, the mean copy number of the elements, $\Lambda$, is also correlated between the species ($r = 0.951, P < 0.05$). Since $\Lambda$ has no direct association with the occupancy distribution, this correlation supports the view that the processes determining copy number and occupancy per site act differently on different elements and that these processes act similarly on an element regardless of the species in which it resides.

## Occurrence of Transposable Elements in Other Species

To examine the occurrence of these mobile sequences in other species, clones were hybridized to filters containing DNAs from 15 species of the *obscura* group and *D. melanogaster*. Three of the elements (DA21, DA60, and DA70) gave very similar profiles. An example of this pattern is provided in figure 2A. These probes hybridized strongly to much of the genome of all species except *D. melanogaster*. Figure 2A agrees well with a genetic distance tree of these species (Lakovaara and Suara 1982). The

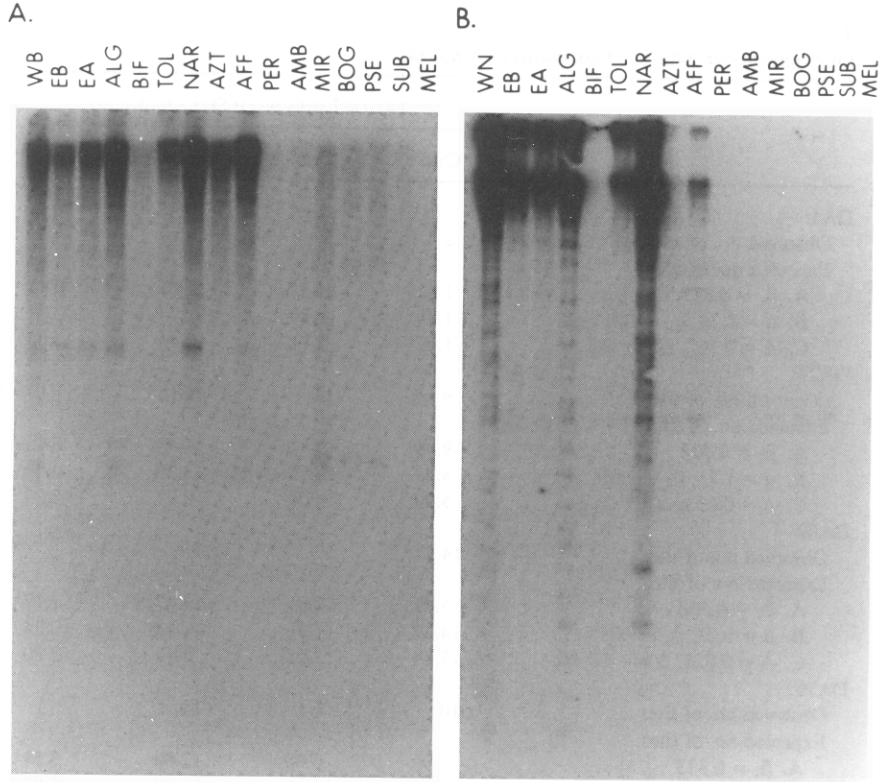


FIG. 2.—Autoradiographs of filters containing genomic DNAs digested with *Bam*HI, clones DA60 (panel A) and clone DA1 (panel B) used as probes. The specific designations are as follows: WN = western-northern semispecies of *Drosophila athabasca;* EB = eastern-B semispecies of *D. athabasca;* EA = eastern-A semispecies of *D. athabasca;* ALG = *D. algonquin;* BIF = *D. bifasciata;* TOL = *D. tolteca;* NAR = *D. narragansett;* AZT = *D. azteca;* AFF = *D. affinis;* PER = *D. persimilis;* AMB = *D. ambigua;* MIR = *D. miranda;* BOG = *D. pseudoobscura bogotana;* PSE = *D. pseudoobscura;* SUB = *D. subobscura;* and MEL = *D. melanogaster.*

difference in hybridization between a species and *D. algonquin* is in rough accord with the genetic distance between them.

In contrast to the other probes, DA1 hybridization was restricted to species of the *affinis* subgroup (fig. 2B; also see Lakovaara and Suara 1982).

## Discussion

### Number of Transposable Elements in Genome

The proportion of *Drosophila algonquin* clones that clearly carry transposable sequences is 17% ($0.26 * {}^{10}/_{15}$). This is similar to estimates of 21% from *D. melanogaster* studies in which randomly picked clones, rather than clones detected by a repetitive-DNA screen, were examined (Young 1979; Dowsett and Young 1982). However, the differences observed between *D. algonquin* and *D. melanogaster* transposable-element copy numbers cannot be an artifact of differences in the selection procedure. Young (1979) examined 80 random clones and found 18 that clearly carried transposable sequences. Of 17 examined closely, all gave unique in situ hybridization patterns. The highest genomic copy number observed was 111, while the mean was 31. If *D. mel-*

*anogaster* carries transposable-element families of high copy number as does *D. algonquin,* then Young should have found about 14 clones (0.17 * 80) fitting the description of the elements observed in *D. algonquin.* Similarly, Dowsett and Young (1982) did Southern hybridizations on 94 random clones to genomic DNA of *D. melanogaster* and found 26 that appeared to carry transposable elements. Of 10 that were examined via in situ hybridization to salivary-gland chromosomes, all appeared different, and none showed more than 60 copies in the genome. Clearly, if *D. melanogaster* carries several elements having a copy number as high as that of elements found in *D. algonquin,* they would have been observed.

When the formulas of Engels (1981), which assume that all families are equally likely, are applied, four families observed among 10 clones leads to an estimate of 4.3 (±0.7) families in the *D. algonquin* genome. However, because of variation in copy number, the assumption of equal sampling is not met. Thus, a small number of additional high-copy-number sequences may occur in the genome. Moreover, several studies have reported that some of the *D. melanogaster* elements occur in low copy number in *D. affinis* or related species (Martin et al. 1983; Lansman et al. 1985; Stacey et al. 1986). Nevertheless, the transposable portion of the genome of *D. algonquin* is clearly dominated by a small number of high-copy-number-element families.

## Importance of High Occupancies per Site

The site-occupancy estimates from *D. algonquin* and *D. affinis* are much higher than those reported for elements in *D. melanogaster.* The data from two studies of X chromosomes from natural populations of *D. melanogaster* are shown in table 2. Values of $Q$ are higher—and both estimates of $\beta$ are much reduced in *D. algonquin* and *D. affinis*—relative to those for *D. melanogaster.* The models of Charlesworth and Charlesworth (1983) and Langley et al. (1983) provide good fits to the observed occupancy distribution for transposable elements in *D. algonquin* (table 3) and *D. melanogaster* (Charlesworth and Charlesworth 1983; Kaplan and Brookfield 1983; Leigh Brown and Moss 1987). Thus, different rates of deletion—and, possibly, transposition—provide an explanation for the site-occupancy contrast that is consistent with the data.

An additional consideration, raised by Kaplan and Brookfield (1983), is that elevated occupancies per site may arise from limited cytological resolution. In this view, which assumes that no insertion occurs more than once in the sample, the region of the genome under investigation is divided into a number of small intervals of size $\Delta$, the length of a cytologically distinguishable region labeled by hybridization. The inverse of $\Delta$, an estimate of the number of cytologically distinguishable sites, can be compared with the observed number of distinguishable sites. For the data of Montgomery and Langley (1983) on *D. melanogaster,* this model fits the data well (Kaplan and Brookfield 1983). For *D. algonquin,* the expected distributions and $\Delta$ values are shown in table 3 (model C). The model is rejected on two levels. First, all $\Delta$ values are considerably less than the actual resolution of at least 53 distinguishable sites. Second, the expected distribution of occupancies fit the data poorly, as shown by the $\chi^2$ values.

An alternative explanation for the high occupancies per site in *D. algonquin* and *D. affinis* is that some sites have a very high probability of receiving an insertion or a very low probability of loss (either via deletion or via selection). If this is the case, then *D. melanogaster* would seem to be free of such hot spots. The hot-spot model

and the models of Charlesworth and Charlesworth (1983) and Langley et al. (1983), which assume that all sites are equally subject to insertion and deletion, lead to different predictions when comparisons are made between closely related species. Because the hot-spot model requires that the basis for hot spots be genetically encoded in some sense, similar species are expected to share hot spots. However, if all sites are subject to the same forces, then frequencies at all sites wll be continually changing and different species should have their high-occupancy sites in different locations.

To test whether sites of high occupancy within species could be explained by sites shared between species, the hybridization data from *D. algonquin* and *D. affinis* were put into the form of a similarity matrix. The locations of insertions in a particular combination of line and clone were represented as a vector of 1's and 0's—for presences and absences, respectively—at all cytologically distinguishable sites. Pearson's product-moment correlation coefficient was used to indicate similarity between two vectors. A matrix of dimension 20 contained the correlation coefficients between all pairs of vectors from the five lines and four clones studied. In fact, two matrices were constructed—one for the data from chromosome arm BL and one for chromosome arm BS.

Mantel (1967) developed a test for the presence of an overall relationship between a data matrix of similarities and a matrix of identical dimension based on some hypothesis of structure in the data matrix. The test is straightforward: a coefficient of association between the two matrices, $Z$ (a sum of cell-by-cell cross-products) is calculated; an empirical distribution of this statistic is determined by a repeated process of random permutation of rows and columns of one matrix and recalculation of $Z$ for each permutation; and, finally, the observed value of $Z$ is compared with the empirical null distribution.

A particular hypothesis of structure in the matrix from *D. algonquin* and *D. affinis* can be described in a matrix with 1's in cells corresponding to locations in the data matrix where high values are expected and 0's in cells corresponding to expectations of lower values. The hypothesis that sites are shared between lines regardless of species is portrayed as a matrix with 1's in those cells corresponding to correlations between data from lines hybridized with the same clone, regardless of what species that line came from (table 4, matrix 1). The hypothesis that high occupancies arise independently within species is depicted by a matrix with 1's only where correlations are between data from two lines of the same species probed with the same clone (table 4, matrix 2).

For each test the empirical distribution of $Z$ was formed from 249 random permutations of the test matrix. If the true value was more extreme than the results of all permutations, the probability ($P$) of obtaining the observed results by chance equaled either 0.004 or 0.996 ($^1/_{250}$ or $^{249}/_{250}$). For a two-tailed test, the chances of obtaining a value equal to or greater than the observed value must be $>0.975$ or $<0.025$ to be significant at $\alpha = 0.05$. All the tests were made using the program NEWMAN3R in the statistical package RPACKAGE (Legendre 1985).

Both similarity matrices had highly significant associations with both test matrices of table 4. In each case the actual Mantel statistic was more extreme than all permutations ($P = 0.996$). Since the two test matrices share the locations of many 1's, a more sensitive test is required.

Smouse et al. (1987) have extended the Mantel test to allow the evaluation of competing, overlapping hypotheses. Their method makes use of the partial correlation coefficient between two matrices calculated while holding other matrices constant. As before, the actual value is contrasted with a null distribution from many random

**Table 4**
**Matrices for Mantel Tests**

A. Matrix 1

| Line and Probe | ALli DA1 | ALbi DA1 | ALmc DA1 | AFho DA1 | AFli DA1 | ALli DA21 | ALbi DA21 | ALmc DA21 | AFho DA21 | AFli DA21 |
|---|---|---|---|---|---|---|---|---|---|---|
| ALli DA1 | | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| ALbi DA1 | | | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| ALmc DA1 | | | | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| AFho DA1 | | | | | 1 | 0 | 0 | 0 | 0 | 0 |
| AFli DA1 | | | | | | 0 | 0 | 0 | 0 | 0 |
| ALli DA21 | | | | | | | 1 | 1 | 1 | 1 |
| ALbi DA21 | | | | | | | | 1 | 1 | 1 |
| ALmc DA21 | | | | | | | | | 1 | 1 |
| AFho DA21 | | | | | | | | | | 1 |
| AFli DA21 | | | | | | | | | | |

B. Matrix 2

| Line and Probe | ALli DA1 | ALbi DA1 | ALmc DA1 | AFho DA1 | AFli DA1 | ALli DA21 | ALbi DA21 | ALmc DA21 | AFho DA21 | AFli DA21 |
|---|---|---|---|---|---|---|---|---|---|---|
| ALli DA1 | | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALbi DA1 | | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALmc DA1 | | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AFho DA1 | | | | | 1 | 0 | 0 | 0 | 0 | 0 |
| AFli DA1 | | | | | | 0 | 0 | 0 | 0 | 0 |
| ALli DA21 | | | | | | | 1 | 1 | 0 | 0 |
| ALbi DA21 | | | | | | | | 1 | 0 | 0 |
| ALmc DA21 | | | | | | | | | 0 | 0 |
| AFho DA21 | | | | | | | | | | 1 |
| AFli DA21 | | | | | | | | | | |

NOTE.—The upper-left-hand portion of the matrices are shown. Not shown are those cells at intersections involving clones DA60 and DA70. For cells not shown, the position of 1's and 0's follows the given pattern.

permutations. This allows a test of association between a data matrix and that component of a test matrix that is independent of similarity with other test matrices.

According to the method of Smouse et al. (1987), the association between the actual data and test matrix 2 was examined while holding test matrix 1 constant. The results were identical and significant for the data from both chromosomes ($P = 0.996$). The results do not support the hot-spot hypothesis. In effect, matrix 2 fits the actual data very well, and this does not depend on the similarity between matrix 1 and matrix 2. Thus, interspecific similarity cannot explain high occupancies within species.

When the association between the data and matrix 2 was tested while holding matrix 1 constant, the results differed between the two data sets. For chromosome arm BS the partial correlation fell in the high tail of the distribution and was significant at the 0.05 level ($P = 0.984$). Matrix 1 appears to have similarity to the data from chromosome arm BS that does not depend on matrix 2, suggesting some additional interspecific pattern. However this was not observed with the data from chromosome BL, for which the partial correlation was lower than all but eight values from the null distribution ($P = {}^{8}/_{250} = 0.032$).

Concerted Evolution and the Age of Transposable Elements

Hybridization of the four transposable-element probes of *D. algonquin* to DNA of 15 species of the *obscura* group reveals three (DA21, DA60, and DA70) that hybridiz across all these species and one (DA1) that hybridizes only in the *affinis* subgroup. I contrast, *melanogaster* elements exhibit a pattern of occurrence across species that often incongruent with the phylogeny of those species (Dowsett 1983; Martin et a 1983; Daniels et al. 1986; Stacey et al. 1986). The data presented here show no evidenc for either loss of elements or horizontal transfer between species. Furthermore, th pattern of decreasing hybridization with increasing genetic distance from *D. algonqui* is exactly the expected pattern if these element families are undergoing concerte evolution. Different copies of an element family taken from within a species resemb each other more than they resemble copies from another species.

## Acknowledgments

This work benefited from discussions with W. Eanes, D. Houle, and J. Ajioka. am indebted to M. Burgman for his help in application of the Mantel tests. Speci thanks is due to W. M. Fitch for providing many useful comments. The research wa supported by National Science Foundation dissertation improvement grant BS 8401798 to J. Hey under the sponsorship of W. Eanes and by National Scienc Foundation grant BSR 8402967 to W. Eanes. This is contribution 696 from Graduat Studies in Ecology and Evolution at SUNY–Stony Brook.

LITERATURE CITED

ANANIEV, E. V., V. E. BARSKY, Y. V. ILYIN, and M. V. RYZIE. 1984. The arrangement transposable elements in the polytene chromosomes of *Drosophila melanogaster*. Chr mosoma **90**:366–377.

BELYAEVA, E. SP., E. V. ANANIEV, and V. A. GVOZDEV. 1984. Distribution of mobile dispers genes (mdg-1 and mdg-3) in the chromosomes of *Drosophila melanogaster*. Chromoson **90**:16–19.

BINGHAM, P. M., R. LEVIS, and G. M. RUBIN. 1981. Cloning of DNA sequences from the *whi* locus of *Drosophila melanogaster* by a novel and general method. Cell **25**:693–704.

CHARLESWORTH, B., and D. CHARLESWORTH. 1983. The population dynamics of transposab elements. Genet. Res. **42**:1–27.

DANIELS, S. B., and L. D. STRAUSBAUGH. 1986. The distribution of P element sequences *Drosophila:* the *willistoni* and *saltans* species groups. J. Mol. Evol. **23**:138–148.

DOWSETT, A. P. 1983. Closely related species of *Drosophila* can contain different libraries middle repetitive DNA sequences. Chromosoma **88**:104–108.

DOWSETT, A. P., and M. W. YOUNG. 1982. Differing levels of dispersed repetitive DNA amoi closely related species of *Drosophila*. Proc. Natl. Acad. Sci. USA **79**:4570–4574.

ENGELS, W. R. 1981. On estimating the total number of genes of a given kind in the genon and similar problems. J. Math. Biol. **11**:45–50.

FINNEGAN, D. J., and D. H. FAWCETT. 1985. Transposable elements in *Drosophila melanogast* Pp. 1–62 *in* N. MACLEAN, ed. Oxford surveys on eukaryotic genes. Vol **3**. Oxford Universi Press, Oxford.

FRISCHAUF, A.-M., H. LEHRACH, A. POUSTKA, and N. MURRAY. 1983. Lambda replaceme vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.

HELMS, C., M. Y. GRAHAM, J. E. DUTCHIK, and M. V. OLSON. 1985. A new method f purifying lambda DNA from phage lysates. DNA **4**:39–49.

HEY, J., and D. HOULE. 1985. Rearing *Drosophila athabasca*. Drosophila Inf. Serv. **61**:192–193

KAPLAN, N. L., and J. F. Y. BROOKFIELD. 1983. Transposable elements in mendelian pop lations. III. Statistical results. Genetics **104**:485–495.

LAKOVAARA, S., and A. SAURA. 1982. Evolution and speciation in the *Drosophila obscura* group. Pp. 2–60 *in* M. ASHBURNER, H. L. CARSON, and J. N. THOMPSON, JR., eds. The genetics and biology of *Drosophila.* Vol. **3B**. Academic Press, New York.

LANGLEY, C. H., J. F. Y. BROOKFIELD, and N. KAPLAN. 1983. Transposable elements in mendelian populations. I. A theory. Genetics **104**:457–471.

LANSMAN, R. A., S. N. STACEY, T. A. GRIGLIATTI, and H. W. BROCK. 1985. Sequences homologous to the P mobile element of *Drosophila melanogaster* are widely distributed in the subgenus Sophophora. Nature **318**:561–563.

LEGENDRE, P. 1985. RPACKAGE: the R package for multivariate data analysis. Départment de Sciences Biologiques, Universite de Montreal, Quebec.

LEIGH BROWN, A. J., and J. E. MOSS. 1987. Transposition of the *I* element and *copia* in a natural population of *Drosophila melanogaster.* Genet. Res. **49**:121–128.

MCGINNIS, W., A. W. SHERMOEN, and S. K. BECKENDORF. 1983. A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. Cell **34**:75–84.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MANNING, J. E., C. W. SCHMID, and N. DAVIDSON. 1975. Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome. Cell **4**:141–155.

MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27**:209–220.

MARTIN, G., D. WIERNASZ, and P. SCHEDL. Evolution of *Drosophila* repetitive-dispersed DNA. J. Mol. Evol. **19**:203–213.

MILLER, D. D. 1939. Structure and variation of the chromosomes of *Drosophila algonquin.* Genetics **24**:699–708.

———. 1977. Salivary gland chromosome variation in the *Drosophila affinis* subgroup. VI. Comparison of X, B, and E chromosomal patterns in *Drosophila athabasca* and five related species. J. Hered. **68**:105–113.

MONTGOMERY, E., B. CHARLESWORTH, and C. H. LANGLEY. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster.* Genet. Res. **49**:31–41.

MONTGOMERY, E. A., and C. H. LANGLEY. 1983. Transposable elements in mendelian populations. II. Distribution of three *copia*-like elements in a natural population of *Drosophila melanogaster.* Genetics **104**:473–483.

RONSSERAY, S., and D. ANXOLABEHERE. 1986. Chromosomal distribution of P and I transposable elements in a natural population of *Drosophila melanogaster.* Chromosoma **94**:433–440.

SMOUSE, P. E., J. C. LONG, and R. R. SOKAL. 1987. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35**:627–632.

STACEY, S. N., R. A. LANSMAN, H. W. BROCK, and T. A. GRIGLIATTI. 1986. Distribution and conservation of mobile elements in the genus *Drosophila.* Mol. Biol. Evol. **3**:522–534.

STROBEL, E., P. DUNSMUIR, and G. M. RUBIN. 1979. Polymorphisms in the chromosomal locations of elements of the 412, *copia* and 297 dispersed repeated gene families in *Drosophila.* Cell **17**:429–439.

YOUNG, M. W. 1979. Middle repetitive DNA: a fluid component of the *Drosophila* genome. Proc. Natl. Acad. Sci. USA **76**:6274–6278.

WALTER M. FITCH, reviewing editor

Received July 6, 1988; revision received July 27, 1988