Proc. Natl. Acad. Sci. USA
Vol. 76, No. 12, pp. 6274–6278, December 1979
Biochemistry

Middle repetitive DNA: A fluid component of the *Drosophila* genome

(repetitive DNA/DNA rearrangements/genome organization/*Drosophila melanogaster*)

MICHAEL W. YOUNG

The Rockefeller University, 1230 York Avenue, New York, New York 10021

Communicated by Norton D. Zinder, September 27, 1979

ABSTRACT **Most of the middle repetitive DNA of *Drosophila melanogaster* appears to be organized into families of 10–100 repeated elements that are found at scattered locations in the chromosome arms and occupy new chromosomal positions as populations of *D. melanogaster* diverge. These "nomadic" DNA segments can be identified by an analysis of cDm plasmids, hybrids of ColE1 and segments of randomly sheared *D. melanogaster* DNA. Eighty cDm plasmids were withdrawn, at random, from a library of approximately 17,000 cDm clones. Fifty-seven of these seem to contain either DNA that is not repeated in the *D. melanogaster* genome or DNA that has a low repetition frequency. The remaining 23 cDm plasmids contain repetitive sequences. Seventeen of these 23 plasmids contain repetitive sequences that are demonstrably scattered to many chromosomal sites that can be mapped in two *D. melanogaster* strains, g-1 and g-X11. The repeated elements hybridizing with each of the different Dm segments are at quite different chromosomal locations in these two strains. However, the size of each family of repeated sequences remains fairly constant in both strains. It is proposed that the number of elements in each family has been fixed by selection.**

The genome of *Drosophila melanogaster* can be broadly separated into three components by reassociation kinetics. The largest component is nonrepetitious DNA. The remainder, one-third of the nuclear DNA, is reiterated and can be further subdivided into sequences that are either highly repetitive, with an average reiteration frequency of about 24,000, or middle repetitive, having an average reiteration frequency of between 35 and 100 (1, 2). These two classes of repetitive DNA compose 19% (3) and 16–17% (calculated from refs. 2 and 3) of the genomic DNA, respectively. Recently, eight families of middle repetitive genes have been described. Each is composed of elements that are arranged as dispersed rather than tandem repeats. Five of these families code for abundant poly(A)-containing RNAs, which allowed their initial isolation (4–11). The remaining three families were recognized in a "chromosomal walk" through 360 kilobases (kb) of DNA from the 87E region of the *Drosophila* genome (12). A remarkable property that is shared by all eight families is that quite different patterns of dispersion are seen when the chromosomal arrangements of these repeated sequences are examined in genetically isolated laboratory stocks or noninterbreeding wild-type populations of *D. melanogaster* (7–12).

From the data presented in this paper we can conclude that the striking rearrangement of the elements of these repeated gene families is not a curiosity limited to a minute fraction of the *Drosophila* genome; rather, variability of position is a property common to most of the middle repetitive DNA. We also present evidence that the number of repeating elements in each family is maintained notwithstanding this variability of position. To simplify this presentation, DNA segments that seem to wander to new chromosomal locations as *Drosophila* populations diverge will be referred to as "nomadic."

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.

MATERIALS AND METHODS

Nucleic Acid Preparation. The cDm plasmids consist of segments of randomly sheared *Drosophila* DNA (Dm segments), strain Oregon R, inserted into the colicinogenic plasmid ColE1. The cDm plasmids used in this study were randomly chosen from a library of approximately 17,000 hybrid plasmids. The construction of this library will be described in detail elsewhere (D. J. Finnegan, G. M. Rubin, D. J. Bower, and D. S. Hogness, unpublished). Briefly, Dm segments were prepared for cloning by shear breakage of Oregon R embryonic nuclear DNA and inserted at the restriction endonuclease *Eco*RI site of ColE1 (13) by the poly(dA)·poly(dT) method of Wensink *et al.* (14). The resulting hybrid DNAs were then cloned by transformation of *Escherichia coli* K-12, strain HB101, to colicin E1 immunity (15). Plasmid DNAs were isolated as described by Finnegan, Rubin, Bower, and Hogness (personal communication). Plasmid DNAs were transcribed *in vitro* with *E. coli* RNA polymerase, as described by Wensink *et al.* (14). Restriction endonuclease digests and agarose gel electrophoreses were carried out as described by Finnegan *et al.* (6). The *D. melanogaster* Oregon R embryonic DNA used in RNA·DNA filter hybridizations was prepared according to the method of Laird and McCarthy (16) with the exception that RNA was removed from these DNA preparations by sifting through a column of controlled-pore glass beads (Electro-Nucleonics, Fairfield, NJ, CPG-10).

Hybridization Procedures. Hybridizations of [^{32}P]cRNA to denatured restriction fragments that had been transferred from agarose gels to nitrocellulose strips by Southern's blotting procedure (17) were carried out as described (5). The method that was used to hybridize [^{3}H]cRNA probes to polytene chromosomes has also been described (14).

Enzymes. Restriction endonucleases were obtained from Bethesda Research Laboratories and New England BioLabs. *E. coli* RNA polymerase was a gift of P. Model.

***D. melanogaster* Strains.** The strains g-1, g-X11, and Oregon R came from the laboratory of David Hogness. The genotype of g-1 is $gt^1\ w^a$;+ and that of g-X11 is $y\ gt^{x11}/FM6$;+. Oregon R is wild type. Our strain g-1 can be traced to the Pasadena stock collection (California Institute of Technology), where it has been maintained as "$gt\ w^a$" since about 1930. Although strain g-X11 has a more complex history, its construction was unrelated to that of g-1. Consequently, we can assume that these strains have been separated for at least 50 years. Strains g-1 and g-X11 are easily interbred and the banding patterns of their polytene chromosomes are homosequential with the exception of a multiply inverted X chromosome balancer (*FM6*) carried in the g-X11 stock. Flies were grown on standard cornmeal agar at 25°C.

Physical and Biological Containment. Where appropriate, P2, EK1 conditions, as described by the National Institutes of

Abbreviations: kb, kilobase; Dm, *D. melanogaster* DNA segments; cDm, Dm segments inserted into colicinogenic plasmid ColE1.

Health Guidelines for Recombinant DNA Research, were used.

RESULTS

Identification of cDm Plasmids that Carry Middle Repetitive DNA. Eighty colicin-resistant colonies, each carrying an independently cloned ColE1–*Drosophila* DNA hybrid plasmid (cDm plasmid) were selected at random from a library of several thousand clones. cDm plasmids carrying middle repetitive DNA can be identified within this collection by a general procedure in which cDm plasmid sequences are hybridized to total genomic DNA. In this procedure, the genomic DNA is prepared in two forms. For initial analysis, embryonic DNA is digested to completion with a restriction endonuclease, in these experiments *Eco*RI, and the resulting fragments are separated according to length by agarose gel electrophoresis. After denaturation and transfer of the fragment strands to a nitrocellulose membrane by Southern's blotting technique (17), they are hybridized with a ^{32}P-labeled probe prepared by *in vitro* transcription of cDm plasmid DNA. As an illustration of this procedure, the results of 3 of 80 hybridizations, each involving total genomic DNA and a single cDm plasmid sequence, are shown in Fig. 1. cDm2026 is a hybrid plasmid that carries 14 kb of Drosophila DNA. Fig. 1, lane a shows that cDm2026 sequences hybridize weakly to restriction fragments of only four or five lengths in *Eco*RI-digested total genomic DNA. In contrast, both cDm2017, which carries only 6.5 kb of *Drosophila* DNA, and cDm2015 (11.5 kb long) hybridize intensely to the genomic digest. Further, cDm2015 shows at least 12 bands of hybridization that represent a total length of DNA (>40 kb) greatly exceeding that of the *Drosophila* sequences cloned in cDm2015. Clearly, sequences found in Dm2017 and Dm2015 are more abundantly represented in the *Drosophila* genome than are those found in Dm2026. As a result of this analysis we conclude that Dm2017 and Dm2015 carry repeated DNA, whereas the sequences found in Dm2026 may be entirely nonrepetitive.

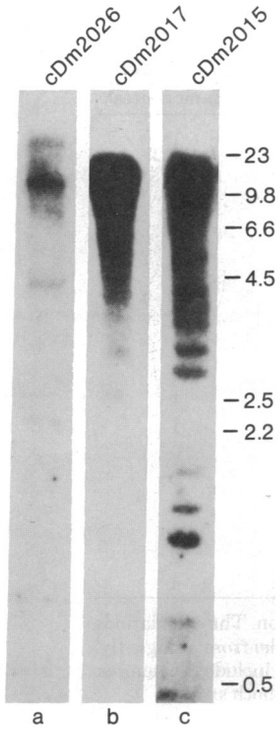

FIG. 1. Comparison of the *Eco*RI restriction fragments from total genomic DNA hybridizing with Dm2026, Dm2017, and Dm2015 sequences. *D. melanogaster* (Oregon R) embryonic nuclear DNA was digested with *Eco*RI and the resulting fragments were separated in a 0.7% agarose gel. Lane a is the autoradiograph obtained when the fragments were hybridized with cDm2026 [^{32}P]-cRNA according to Southern's (17) method. Lane b is the result obtained when hybridization was to cDm2017 [^{32}P]cRNA. For lane c the probe was the ^{32}P-labeled cDm2015 sequence. The lengths, in kb, of the *Hin*dIII fragments of bacteriophage λ (18) are presented to the right.

As a complement to the experiments described above, the salivary gland polytene chromosomes of two laboratory stocks of *D. melanogaster*, g-1 and g-X11, were prepared for *in situ* hybridization to each of the 80 cDm plasmids. Fig. 2 *a* and *b* shows that [^{3}H]cRNA copied from cDm2026 hybridizes to a single chromosomal position, region 12 D-F of the X chromosome, in both strains. No other site, either in the chromosome arms or within the chromocenter, is labeled in g-1 or g-X11 with this probe. Sequences found in cDm2017 can be found at two

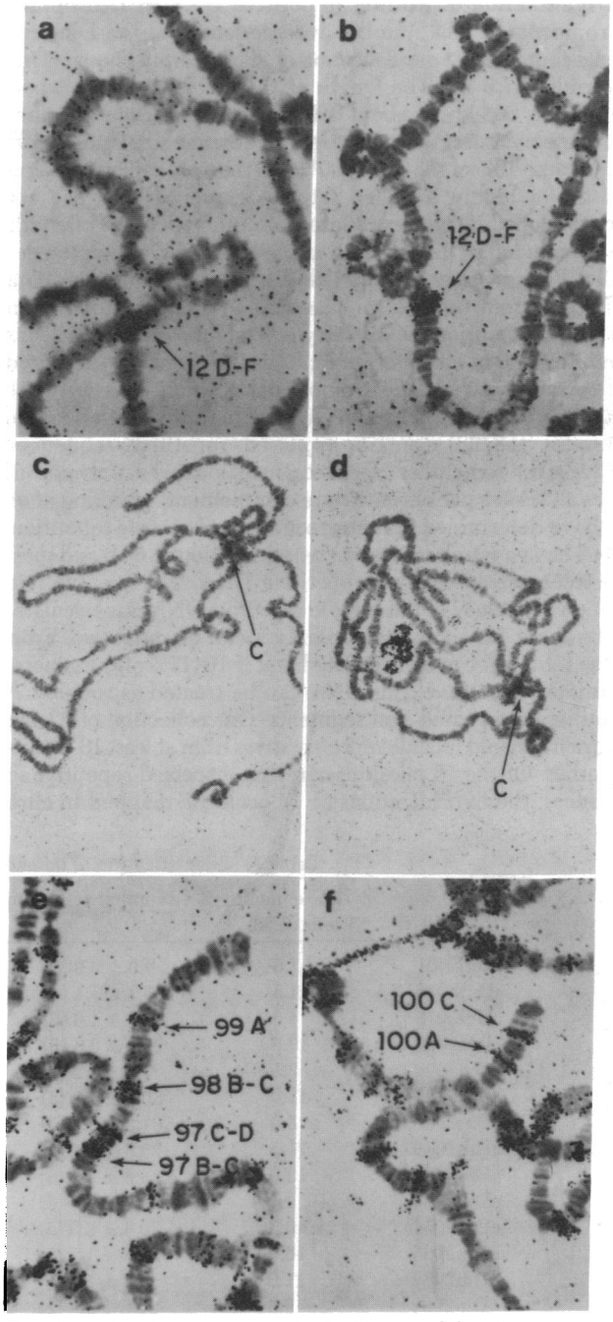

FIG. 2. Mapping the Dm2026, Dm2017, and Dm2015 sites in polytene chromosomes by *in situ* hybridization. [^{3}H]cRNA transcribed from cDm2026 was hybridized with polytene chromosomes from g-1 (*a*) and g-X11 (*b*). Dm2017 sequences were hybridized to g-1 (*c*) and g-X11 (*d*), and Dm2015 sequences were hybridized to g-1 (*e*) and g-X11 (*f*). The number and letter designations appearing in each photomicrograph refer to the chromosomal regions mapped by Bridges (19). In *c* and *d*, C indicates chromocenter. *d* also shows labeling of the nucleolus. Exposure time was 60 days for each experiment.

positions after *in situ* hybridization, the chromocenter and nucleolus (Fig. 2 *c* and *d*). Identical labeling patterns are observed after hybridization of cDm2017 sequences to either strain g-1 or g-X11. Apparently, cDm2017 contains sequences that are arranged as a clustered repeat at one or both of these chromosomal positions. Fig. 2 *e* and *f* shows the hybridization pattern of cDm2015 sequences. cDm2015 contains a segment of DNA that is repeated in a dispersed fashion. Further, a portion of Dm2015 is nomadic; the arrangement of this dispersed repetitious DNA is different in g-1 and g-X11.

Organization of Middle Repetitive DNA in *Drosophila* Chromosomes. The number-average length of the Dm segments contained in our collection of 80 randomly chosen cDm plasmids is 7.6 kb. The length of these Dm segments when summed is 608 kb, or about 0.4% of the *Drosophila* genome. For these measurements the effective genome size is 134×10^3 kb, because 19% of the 165×10^3 kb in the total haploid genome consists of highly repeated satellite sequences that are not usually cloned by our procedures (6). Although all 80 plasmids were included in the two-step analysis outlined above, the detailed listings presented in Table 1 represent only the data collected from those cDm plasmids that appear to contain dispersed repetitious DNA. Fifty-seven of the 80 Dm segments seem to contain either no repetitious DNA or DNA repeated a few times at a single chromosomal position. Twenty-three plasmids carry repetitive DNA. The majority of these cDm plasmids (18/23) contain dispersed repetitive sequences, whereas the remainder carry sequences located at only one (3) or two (2) sites per chromosome complement. Manning *et al.* (1) have determined that most segments of middle repetitious DNA have a number average length of about 5.6 kb and these are interspersed with nonrepetitive DNA. The average reiteration frequency of this middle repetitive DNA was calculated to be about 72 with an uncertainty factor of about two. From these figures it can be calculated that if 16–17% of the genome is middle repetitive, this DNA can be treated as roughly 70 families of repeated Dm segments. Our collection of 23 Dm segments should include representatives from at least 18 of these families. Of the 18 plasmids carrying dispersed repetitive sequences, the hybridizations of 17 could be mapped in chromosome arm 3R in both strain g-1 and g-X11 (Table 1). By comparing these maps, we found that all 17 contain a nomadic DNA sequence. From a casual inspection of our chromosome preparations we conclude that for each nomadic Dm segment the degree of polymorphism exhibited by chromosome arm 3R is similar to that found in the remaining chromosome arms.

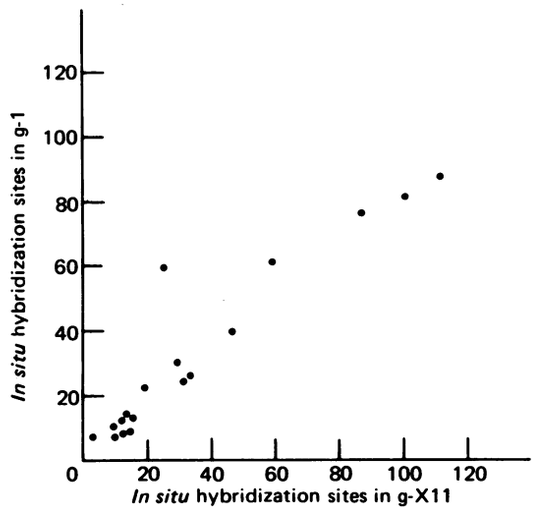

FIG. 3. Size distribution of nomadic DNA families. The number of *in situ* hybridization sites for each of 18 families in strain g-1 is plotted as a function of number of *in situ* hybridization sites in strain g-X11. Hybridization to the chromocenter was excluded from this analysis. The data used in forming the 18 points are presented in Table 1.

Several additional findings are noteworthy. cDm plasmids containing dispersed repeated sequences hybridized to an average (arithmetic mean) of 32.6 polytene chromosomal sites in strain g-1 and 29.4 in strain g-X11. In g-1 the total range of sites hybridized per Dm segment was 6.6–87.3, and in strain g-X11 it was 2.5–111.2 (Fig. 3).

Table 1 shows that although wide variations in repetition frequency occur among nomadic families, the number of chromosomal sites occupied by a single nomadic sequence is

Table 1. Identification of nomadic DNA in middle repetitive Dm segments

Plasmid	Dm segment length, kb	Number of genomic hybridization sites		Number of 3R hybridization sites		
		g-1	g-X11	g-1	g-X11	Common sites
cDm2001	7.5	9.0 ± 0.9 (6)	14.4 ± 0.6 (5)	1	5	0
cDm2015	11.5	87.3 ± 5.4 (6)	111.2 ± 6.4 (5)	23	25	9
cDm2016	4.5	13.8 ± 1.6 (5)	12.8 ± 1.3 (5)	5	6	3
cDm2027	19.5	23.8 ± 2.5 (6)	31.0 ± 1.4 (5)	4	5	0
cDm2028	3.5	7.8 ± 1.2 (6)	10.8 ± 3.0 (5)	1	1	0
cDm2029	11.0	80.6 ± 4.8 (5)	100.4 ± 1.1 (5)	19	23	9
cDm2041	5.0	61.3 ± 3.8 (3)	59.4 ± 6.1 (5)	14	14	9
cDm2046	11.5	25.8 ± 2.1 (6)	33.0 ± 1.0 (5)	6	10	2
cDm2047	13.0	13.3 ± 2.0 (6)	15.0 ± 2.2 (5)	2	6	2
cDm2054	4.0	9.8 ± 0.8 (5)	8.8 ± 0.5 (5)	2	1	0
cDm2064	5.0	39.4 ± 2.3 (5)	45.8 ± 3.2 (5)	3	16	1
cDm2066	16.5	29.6 ± 1.7 (5)	29.3 ± 1.9 (6)	6	8	1
cDm2067	11.0	6.7 ± 1.4 (6)	2.5 ± 0.6 (6)	1	0	0
cDm2068	9.5	22.0 ± 1.4 (5)	18.8 ± 1.7 (6)	6	3	0
cDm2073	5.0	76.2 ± 4.7 (5)	87.2 ± 4.0 (5)	Not determined		
cDm2074	7.0	7.0 ± 0.8 (5)	10.0 ± 1.4 (5)	3	2	1
cDm2078	5.0	59.0 ± 2.6 (5)	25.6 ± 2.5 (5)	17	8	2
cDm2088	6.5	11.8 ± 1.0 (5)	12.4 ± 2.0 (5)	2	5	1

Number of sites hybridizing *in situ* represents the arithmetic mean and includes the standard deviation. The calculations do not include hybridization to the chromocenter. Each mean was determined by counting several nuclei from a single fly. The number of nuclei counted for each mean is shown in parentheses. Data given for chromosome arm 3R include the number of *in situ* hybridizations in strains g-1 and g-X11, and the *maximal* number of these sites common to both strains.

fairly constant in both strains g-1 and g-X11. This relationship is also depicted in Fig. 3. For example, cDm2015 hybridizes to 87 sites in the chromosome arms of g-1 and to 111 sites in the arms of g-X11. cDm2073 hybridizes to 76 positions in g-1 and 87 positions in g-X11. The repetition frequency of the nomadic sequence cloned in cDm2001 is much lower in both g-1 (9 sites), and g-X11 (14 sites). Similarly, cDm2074 hybridizes to only 7 sites in g-1 and only 10 sites in g-X11. These results are of special interest in light of the data presented in Table 2, in which the complete hybridization maps of three repeated Dm segments are presented. The nomadic sequence in cDm2074, mentioned above, is found in 7 positions in g-1. Only one of these sites can possibly overlap any of the 10 positions to which cDm2074 hybridizes in g-X11. The situation with cDm2054 and cDm2068 is much the same (Table 2). cDm2054 hybridizes, *in situ*, to 8 sites in strain g-1 and 9 sites in g-X11 with a maximum overlap of 2 positions. cDm2068 hybridizes at 23 positions in g-1 and 20 positions in strain g-X11. In this case no more than 4 of these sites are occupied by the nomadic sequence in both strains. The data presented for chromosome arm 3R, Table 1, show that the same degree of polymorphism exists for each of the remaining dispersed repetitive sequences. This tendency for reiteration frequency to be conserved even though most of the repeating elements of each family occupy different positions in noninterbreeding populations could be readily explained if the size of each family is determined by selection.

In both strain g-1 and strain g-X11, most dispersed repeated Dm segments (17/18) hybridize, *in situ*, to the chromocenter as well as to the chromosome arms. Each dispersed repetitive Dm sequence hybridized uniformly to most of the homologous sites in the polytene chromosome arms. This is to be contrasted with the labeling of the chromocenter; the number of grains localized within the chromocenter was generally one to four times that found at a single site in the arms. In four instances, however, the chromocentral hybridizations were 15–20 times the level of a site in the arms (Dm2027, Dm2047, Dm2066, and Dm2088). The chromocentral hybridizations of each of these four cloned sequences were equally intense in strains g-1 and g-X11. A few nomadic sequences may be repeated many times at the chromocenter. Alternatively, some of our cloned Dm segments may include more than one class of repetitive DNA. After *in situ* hybridization, sequences carried by the Dm segments listed in Table 1 either labeled the chromocenter in both strain g-1 and strain g-X11 or they labeled the chromocenter of neither strain.

DISCUSSION

Most of the middle repetitive DNA in *Drosophila* seems to consist of sequences that are scattered about the genome. This is consistent with the work of Manning *et al.* (1), who concluded from their analysis that most of the middle repetitive DNA in *D. melanogaster* is dispersed rather than clustered. Our results indicate that these sequences are also nomadic.

In *D. melanogaster* scattered middle repetitive DNA is, on the average, about 10–20 times longer than the prevalent interspersed repetitive DNA found in many eukaryotes (1). Further, the middle repetitive DNA of many eukaryotes exhibits a broad range of sequence homology (20, 21). In contrast to these results, Wensink (22) has measured the sequence homology among members of all the different middle repetitive DNA classes in *D. melanogaster* strain Oregon R, and describes each class as a collection of repeated elements of very similar nucleotide sequence, with most members of a class differing by only 3–7%. The analyses of three nomadic gene families, 412 (4, 6, 10, 11), *copia*, (5, 6, 10, 11) and 297 (10, 11) also indicate that these middle repetitive DNAs can be thought of as groups of nearly identical elements. The elements of at least two of these families, 412 and *copia*, share another feature. Even though 412 and *copia* sequences show no sequence homology, they are both terminally redundant; 412 elements are 7.3 kb long and carry direct repeats of 0.5 kb, while *copia* elements, about 5 kb long, contain direct repeats of 0.3 kb (6). The data we have presented are consistent with the possibility that most of the middle repetitive DNA of *Drosophila* is organized into elements having properties similar to those of 412 and *copia*.

Repeated sequences with properties that correlate well with those of *copia* and 412 have also been found in *Saccharomyces cerevisiae*. Cameron *et al.* (23) have described a family of about

Table 2. Chromosomal localization of Dm2054, Dm2068, and Dm2074 *in situ* hybridization sites

	X		2L		2R		3L		3R		4	
Plasmid	g-1	g-X11	g-1	g-X11	g-1	g-X11	g-1	g-X11	g-1	g-X11	g-1	g-X11
cDm2054	4E–F	4E–F	21D			42A–B		66B–C	82A		0	0
		10A–B		30C–D	50C		67C		84E–F			
		18C–D						71A–B		87F–88A		
	19B–D	19D–F					78F					
cDm2068		1D	22A			42A–B		61B	82A–B		0	0
		3C		25D–E		57E–F	61D			84E		
	4D	4D–E	28A		59A–B		64C		85A–C			
	5A		30C–D				67D–E		85E–F			
	6E–F			31F–32A			69A–B	69A–B	87A–B			
	7C			38A–B			71A			90C		
		7E–F	39B–C	39B–C				74C	95E–F			
	10B	10B						75C		96F–97A		
	13A–B						76A		100A			
		14B–C						78C–D				
cDm2074		5D	0	0		42B	64A			82B–C	0	0
	10B–C				43A–B			66A	84C–D			
						55F–56A–B		67C–D	85E			
						57A		68B–C	86C	86C		
							70A					
								70D				

In addition to the sites given, each of these Dm sequences is found at the chromocenters of both strains. Dm2068 is unique among our cloned nomadic sequences in that it also hybridizes, lightly, to the nucleoli of these strains.

35 repetitive elements, Ty1, which are each 5.6 kb long and include a noninverted 0.25-kb sequence of another repetitious family, δ, on each end. Some of the members of this family are tandemly repeated or circular or both, but most are dispersed. Ty1 elements, like the nomadic DNAs of *Drosophila*, are found in nonidentical, scattered arrangements in different yeast strains and new sequence alterations have been detected after culturing a single yeast clone for as little as one month. It has been calculated that Ty1 represents about 2% of the DNA in *Saccharomyces*. Two more repetitious sequences, which are found in varied arrangements in different yeast strains, have been detected by Cameron *et al.* (23). Both families contain about 20 members. Together, these three families must make up a significant portion of the repetitive DNA sequences in yeast; the original reassociation kinetics of yeast DNA revealed little or no repetitive DNA (24).

A final parallel to be drawn concerns the expression of poly(A)-containing RNAs in yeast and *Drosophila*. The most abundant of these are transcribed from nomadic DNAs in both systems. In *D. melanogaster*, *copia* RNA makes up 3–4% of the total poly(A)-RNA in cultured cells (5, 6) and is quite abundant in embryos (25). *copia* appears to be the most abundant poly(A)-RNA in cultured *Drosophila* cells (5, 6). T. St. John (see ref. 23) has shown that RNA homologous to Ty1 is one of the most abundant cellular RNAs in yeast. Only the ribosomal RNAs and double-stranded killer RNAs seem to be more highly represented in this organism.

Evidence has been presented that the number of repeated elements forming each nomadic family is much the same in *D. melanogaster* strain g-1 and strain g-X11. This is true even though the chromosomal positions occupied by the elements of a single nomadic family are different in the two strains (Tables 1 and 2 and *Results*). Fig. 3 depicts the tendency for the size of each family to be the same in both strains. For each family the number of *in situ* hybridization sites in g-1 is plotted as a function of number of sites in g-X11. The 18 points that are generated define a line with a slope that is close to one. For each of the 18 Dm segments presented in Table 1, the number of *in situ* hybridization sites in strain g-1 and g-X11 can also be used in a test for covariance. The numbers of hybridizations in each strain are covariant with a correlation coefficient of 0.94.

There is a more direct test that can be used to measure any mechanism controlling family size, or more importantly, the functional significance of these families. Because the chromosomal positions occupied by a family of repeated sequences are relatively stable within a small inbreeding laboratory stock (unpublished data), it should be possible to drastically change the size of one of these families by recombination between strains. For example, one might attempt to construct, by recombination, a new *D. melanogaster* strain from strains g-1 and g-X11. In this case the number of Dm2068 elements occupying the chromosome arms could be reduced from 23 and 20 members to 4 (Table 2). Our ability to form and maintain a strain of this sort will depend upon the importance of the family of repeated sequences and should reveal the strength of any mechanism controlling dosage.

I thank Ruth Weissberger for technical assistance, and Laurel Eckhardt, Peter Model, and Norton D. Zinder for very helpful criticism of the manuscript. This research was supported by the Andre and Bella Meyer Foundation and the National Institutes of Health.

1. Manning, J. E., Schmid, C. W. & Davidson, N. (1975) *Cell* **4**, 141–155.
2. Schachat, F. H. & Hogness, D. S. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 371–381.
3. Brutlag, D., Appels, R., Dennis, E. S. & Peacock, W. J. (1977) *J. Mol. Biol.* **112**, 31–47.
4. Rubin, G. M., Finnegan, D. J. & Hogness, D. S. (1976) *Prog. Nucleic Acids Res. Mol. Biol.* **19**, 221–226.
5. Young, M. W. & Hogness, D. S. (1977) *Eukaryotic Genetic Systems. ICN UCLA Symposia on Molecular and Cellular Biology* (Academic, New York), Vol. 8, pp. 315–331.
6. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant Biol.* **42**, 1053–1063.
7. Ilyin, Y. V., Tchurikov, N. A., Ananiev, E. V., Ryskov, A. P., Yenikolopov, G. N., Limborska, S. A., Maleeva, N. E., Gvozdev, V. A. & Georgiev, G. P. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 959–969.
8. Tchurikov, N. A., Ilyin, Y. V., Ananiev, E. V. & Georgiev, G. P. (1978) *Nucleic Acids Res.* **5**, 2169–2187.
9. Ananiev, E. V., Gvozdev, V. A., Ilyin, Y. V., Tchurikov, N. A. & Georgiev, G. P. (1978) *Chromosoma* **70**, 1–17.
10. Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979) *Cell* **17**, 415–427.
11. Strobel, E., Dunsmuir, P. & Rubin, G. M. (1979) *Cell* **17**, 429–439.
12. Bender, W., Spierer, P. & Hogness, D. S. (1979) *J. Supramol. Struct. Suppl.* **3**, 32 (abstr.).
13. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3455–3459.
14. Wensink, P. C., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) *Cell* **3**, 315–325.
15. Glover, D. M., (1976) *New Techniques in Biophysics and Cell Biology* (Wiley, London), Vol. 3, pp. 135–145.
16. Laird, C. D. & McCarthy, B. J. (1968) *Genetics* **60**, 303–322.
17. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
18. Murray, K. & Murray, N. E. (1975) *J. Mol. Biol.* **98**, 551–564.
19. Bridges, C. B. (1935) *J. Hered.* **26**, 60–64.
20. Britten, R. J. & Davidson, E. H. (1971) *Q. Rev. Biol.* **46**, 111–138.
21. Davidson, E. H. & Britten, R. J. (1973) *Q. Rev. Biol.* **48**, 565–613.
22. Wensink, P. C. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1033–1039.
23. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16**, 739–751.
24. Britten, R. J. & Kohne, D. E. (1968) *Science* **161**, 529–540.
25. Carlson, M. & Brutlag, D. (1978) *Cell* **15**, 733–742.