Nonrandom Distribution of Long Mono- and Dinucleotide Repeats in *Drosophila* Chromosomes: Correlations with Dosage Compensation, Heterochromatin, and Recombination

KY LOWENHAUPT,* ALEXANDER RICH, AND MARY LOU PARDUE

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 9 August 1988/Accepted 7 December 1988

Long stretches of $(dC-dA)_n \cdot (dT-dG)_n$, abbreviated CA/TG, have a distinctive distribution on *Drosophila* chromosomes (M. L. Pardue, K. Lowenhaupt, A. Rich, and A. Nordheim, EMBO J. 6:1781–1789, 1987). The distribution of CA/TG suggests a correlation with the overall transcriptional activity of chromosomal regions and with the ability to undergo meiotic recombination. These correlations are conserved among *Drosophila* species and may indicate one or more chromosomal functions. To test the generality of these findings, we analyzed the distribution of the rest of the six possible mono- and dinucleotide repeats (A/T, C/G, AT/AT, CA/TG, CT/AG, and CG/CG). All but CG/CG were present at significant levels in the genomes of the six *Drosophila* species studied; however, A/T levels were an order of magnitude lower than those of the other sequences. Data base analyses suggested that the same sequences are present in other eucaryotes. Like CA/TG, both CT/AG and C/G showed increased levels on dosage-compensating chromosomes; however, the individual sites clearly differed for each sequence. In contrast, A/T and AT/AT, although present in *Drosophila* DNA, could not be detected in situ in polytene chromosomes. We also used in situ hybridization to analyze the neo-Y chromosome of *Drosophila miranda*, an ancestral autosome that has become attached to the Y chromosome and is now partially heterochromatic. The neo-Y has acquired repeated DNA sequences; we found that the added sequences are as devoid of mono- and dinucleotide repeats as other heterochromatin. The distribution and function of these sequences are likely to result from both their repetitious nature and base contents.

Both hybridization studies and direct DNA sequencing have revealed that eucaryotic nuclear genomes contain significant numbers of long mono- and dinucleotide repeats. In contrast, these long repeats are not found in procaryotic genomes (15). One especially interesting feature of these simple repeats is their facility for adopting conformations other than the canonical B-DNA structure. Although mono- and dinucleotide repeats are not unique in their ability to form non-B-DNA structures, the very regularity of the repeat sequence, combined with the difference in physical properties of each base, makes these sequences more apt to be structurally labile than so-called random DNA sequences. Such conformations may have biological relevance, as seen in the bending of kinetoplast DNA by short A/T tracts (20). These simple repeating DNAs may play a role in both genomic structure and function. Thus, it is of interest to determine their distribution within a eucaryotic genome. *Drosophila* is a useful organism for such a study because the polytene chromosomes provide unique resolution for sequence mapping. In addition, a wealth of knowledge about the evolution of the karyotype within the genus (30) provides several natural tests for the possible significance of sequence distributions.

In a previous study we reported that long CA/TG repeats have a distinctive distribution in the chromosomes of *Drosophila* spp. In situ hybridization showed CA/TG repeats throughout the euchromatic regions of the chromosomes but none of these were detected in the heterochromatic regions (29). The lack of hybridization in heterochromatic regions was surprising because these regions hybridize strongly with many different families of repeated DNA, including most mobile elements. The euchromatic sites of CA/TG show

considerable evolutionary stability; locations of CA/TG hybridization are very similar in two *Drosophila melanogaster* stocks that differ extensively in the sites of all mobile elements that have been studied (11, 27, 38). A very striking feature of the hybridization pattern is the high concentration of CA/TG sequences on the X chromosome. In each of the six species of *Drosophila* that we have studied, the X chromosome has at least twice as many long CA/TG repeats as does any autosome. The only significant evolutionary changes in the distribution of CA/TG repeats were seen in *D. pseudoobscura* and *D. miranda*. In both species, an autosome of the ancestral karyotype has become attached to the X chromosome to create a metacentric chromosome. This new arm of the X chromosome has acquired the ability to dosage compensate so that, in the hemizygous male, genetic activity is increased to equal the activity of the two X chromosomes of the female (1, 25). This new arm of the X chromosome has also acquired the higher level of CA/TG sequences typical of the X chromosome (18, 29). In *D. miranda*, a second ancestral autosome has become attached to the Y chromosome (24). The autosomal portion of this neo-Y has become partially heterochromatic. The autosomal partner of the neo-Y (now called the X₂ because of its pattern of inheritance) shows regions that can dosage compensate, presumably because heterochromatization of the neo-Y chromosome has made these regions effectively hemizygous in the male (10, 33). The X₂ has also acquired higher levels of CA/TG (29).

Our previous results have shown an intriguing correlation between the distribution of CA/TG and the overall activity of chromosomal regions. CA/TG sites are found throughout the euchromatic, but not heterochromatic, regions of the chromosomes. In addition, CA/TG sequences are at least twice as prevalent in chromosomes that dosage compensate. The

* Corresponding author.

genetic activity of dosage-compensating chromosomes is doubled when these chromosomes are in males. A second correlation is seen between the presence of CA/TG repeats and meiotic recombination; the heterochromatic regions and dot chromosomes of most species, both of which lack CA/TG sites, are also distinguished by lack of meiotic recombination.

We have now determined the amount and distribution of the six possible mono- and dinucleotide repeats in the genomes of several *Drosophila* species. Two of these, CT/AG and C/G, have chromosomal distributions that generally resemble that of CA/TG, although there are distinctive sequence-specific features. On the other hand, A/T and probably AT/AT have quite different chromosomal distributions. We have been unable to find CG/CG in *Drosophila* genomes.

MATERIALS AND METHODS

Abbreviations. Abbreviations used in this paper are as follows: A/T, $(dA)_n \cdot (dT)_n$; C/G, $(dC)_n \cdot (dG)_n$; AT/AT, $(dA-dT)_n \cdot (dA-dT)_n$; CA/TG, $(dC-dA)_n \cdot (dT-dG)_n$; CG/CG, $(dC-dG)_n \cdot (dC-dG)_n$; CT/AG, $(dC-dT)_n \cdot (dA-dG)_n$.

***Drosophila* stocks.** The *D. melanogaster* stocks used were gt-1 and gt-XII (27). The *D. simulans* and *D. hydei* stocks were obtained from W. Gelbart, *D. virilis* was from M. Blumenfeld, *D. pseudoobscura* from R. Lewontin, and *D. miranda* S-204 was from R. Norman.

Computer analysis. The EMBL Nucleotide Sequence Data Library, Release 11.0, was analyzed by using MATCH, a program designed by the National Biomedical Research Foundation. Perfect matches 10 base pairs or longer were counted.

Hybridization probes. Alternating or homopolymers were purchased from Pharmacia and radiolabeled by nick translation. The specific activity of ^3H -labeled polymers was 0.8×10^6 to 3×10^6 cpm/ μg , and 10^5 cpm were used for each slide. ^{32}P -labeled polymers had a specific activity of 2×10^7 to 8×10^7 cpm/ μg and had a single-stranded length of 100 to 200 bases.

DNA isolation and hybridization analysis. Total genomic DNA was isolated from adult heads (chosen because they have less difference in polyteny between males and females than do other body parts), as described by Pardue et al. (29), and bound to dots on nitrocellulose by the method of Baker (4). Alternatively, total genomic DNA was cleaved with *Eco*RI, size separated on a 1% Tris-acetate-EDTA agarose gel, and Southern blotted. Dilutions of polymer over 2 orders of magnitude of concentration, spanning the experimental values, were loaded in each format as standards. In each case, very good linearity existed between amount of polymer loaded and number of counts detected. In addition, there was good agreement between duplicate standard curves, adjusted for specific activity of the probe, and different polymers, except as noted (see Table 1). Nitrocellulose filters were probed with an excess of ^{32}P -labeled polymer. Hybridizations and washes were in $2\times$ SSC ($1\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 68°C for 1.5 h. Hybridization to each dot or to lanes on gels was quantitated with a Betagen Blot Analyser as described by Sullivan et al. (35).

In situ hybridization. In situ hybridization was performed as described by Pardue et al. (29).

RESULTS

***Drosophila* genomes contain significant amounts of long mono- and dinucleotide repeats.** The level of each of the six

possible homopolymer and alternating copolymer repeats in the genomic DNA of several *Drosophila* species was determined by hybridization to filter-bound DNA (Table 1). Five of the six sequences are variously prevalent. The sixth sequence, CG/CG, is very poorly detected by hybridization; however, our results are consistent with other evidence that the *Drosophila* genome does not contain long CG/CG repeats (16, 26; see Discussion).

Different *Drosophila* species show significant differences in their levels of the simple repeating sequences. *D. simulans* contains somewhat less of each of the sequences than do the other species, a result consistent with other studies showing that *D. simulans* has fewer repetitive sequences than other *Drosophila* species (12). *D. melanogaster* has higher levels of simple repeats than *D. simulans*, but slightly lower levels than do the other four species, except in the case of C/G. The level of C/G is very low in *D. hydei* and somewhat depressed in *D. virilis*; however, both of these species have high levels of CA/TG and CT/AG and, relatively speaking, of A/T.

All of the detectable repeating sequences are approximately as prevalent in males as in females. Since the Y chromosome does not replicate in polytene nuclei, its sequence content cannot be confirmed by in situ hybridization; however, the quantitative hybridization suggests that the simple repeats on the Y chromosome are approximately equal to those on the X chromosome. It has recently been reported that the Y chromosome in *D. hydei* contains high levels of CA/TG (18).

The experiments described here cannot distinguish between many shorter sequences and fewer long homopolymer or dinucleotide stretches. A previous analysis of the DNA sequence data base (15) suggested that the size distributions of AT/AT and CG/CG differ from those of CA/TG and CT/AG. We have updated that analysis and included A/T and C/G (Fig. 1). The graphs show that there are, indeed, different distributions for these repeating sequences. CG/CG repeats are found at levels barely higher than random, and no long CG/CG repeats have been encountered. A/T repeats are very prevalent and predominantly ($>98\%$) shorter than 30 base pairs. Several of the longest A/T repeats are from procaryotic sources. AT/AT repeats are also prevalent; $<3\%$ are longer than 30 base pairs. Approximately one-third of all AT/AT repeats were encountered in mitochondrial DNA. In contrast, CA/TG and CT/AG show a distinctly different size distribution, with fewer shorter sequences and a significant number of repeats >30 base pairs and some >100 base pairs. C/G repeats are not as prevalent as CA/TG and CT/AG. Few C/G stretches have been found near the *Drosophila* genes that have been sequenced.

Three of the simple sequence repeats are found at many conserved sites on the euchromatic arms in each *Drosophila* species and are enriched in dosage-compensating chromosomes. In situ hybridization to polytene chromosomes shows that the distributions of CT/AG and C/G fit the general pattern that we have reported for CA/TG; however, each of these sequences has a unique distribution that can be recognized and that is conserved even in distantly related species (Fig. 2). CA/TG is broadly distributed, present to a limited degree in almost every euchromatic chromosomal region, and shows a marked enrichment on the X chromosome. In species with multiple X chromosome arms (Fig. 3), CA/TG sequences are enriched on each of the X arms. CA/TG sequences are never detected in the beta-heterochromatin of the chromocenter or, with the exception of *D. virilis*, on the dot chromosome (the very tiny chromosome seen in most

TABLE 1. Quantitation of mono- and dinucleotide repeats in the genomes of various *Drosophila* species^a

<i>Drosophila</i> species	ng of sequence/μg of genomic DNA (range)					
	CA/TG	CT/AG	C/G	A/T	AT/AT ^b	CG/CG ^c
<i>D. melanogaster</i> gt-1						
Females	0.7 (0.7–0.8)	0.9 (2) ^e	0.6 (0.5–0.7)	0.02 (0.01–0.02)	0.7 (0.5–0.8)	ND ^d
Males	1.1 (0.9–1.2)	0.6 (2)	0.8 (0.8–1.0)	0.03 (0.02–0.04)	0.6 (0.4–0.9)	ND
<i>D. melanogaster</i> gt-X11						
Females	0.7 (0.7–0.8)	0.8 (0.7–0.9)	0.5 (0.4–0.6)	0.02 (0.01–0.02)	0.4 (0.3–0.5)	ND
Males	1.0 (0.9–1.1)	0.7 (3)	0.7 (0.5–0.9)	0.02 (2)	0.6 (0.3–0.7)	ND
<i>D. simulans</i>						
Females	0.5 (1)	0.5 (0.5–0.6)	0.6 (1)	0.01 (1)	0.3 (0.3–0.5)	ND
Males	0.8 (1)	0.4 (0.4–0.5)	0.6 (1)	0.01 (1)	0.1 (2)	ND
<i>D. miranda</i>						
Females	2.0 (1.9–2.5)	1.7 (1.2–2.2)	3.7 (2.8–4.7)	0.02 (0.01–0.03)	0.5 (0.3–0.8)	ND
Males	2.2 (1.9–2.4)	2.0 (1.8–2.2)	3.4 (2.1–4.1)	0.02 (2)	0.4 (0.3–0.7)	ND
<i>D. pseudoobscura</i>						
Females	1.7 (1.6–1.9)	1.8 (3)	2.3 (1.8–2.9)	0.02 (2)	0.6 (0.5–0.9)	ND
Males	1.9 (1.8–2.0)	2.0 (3)	2.0 (1.2–2.7)	0.02 (0.02–0.03)	0.6 (0.4–0.7)	ND
<i>D. hydei</i>						
Females	5.2 (5.0–5.5)	4.1 (4.0–4.3)	0.2 (0.2–0.3)	0.07 (0.04–0.07)	1.1 (0.6–1.2)	ND
Males	6.5 (3)	4.1 (3)	0.4 (0.3–0.5)	0.05 (0.05–0.06)	0.9 (0.4–1.1)	ND
<i>D. virilis</i>						
Females	4.0 (3.9–4.1)	7.0 (5.0–9.0)	0.4 (0.3–0.6)	0.1 (2)	0.7 (0.4–0.9)	ND
Males	5.5 (5.0–6.5)	5.5 (4.0–5.5)	0.4 (0.2–0.5)	0.2 (0.1–0.3)	0.9 (0.5–1.7)	ND

^a DNA was fragmented by using *Eco*RI and either denatured and applied directly to a nitrocellulose filter as a series of dots or separated by agarose gel electrophoresis and transferred to nitrocellulose as a Southern blot. A series of standards containing known dilutions of polymer was applied in the same format in each case, and each filter was probed with ³²P-labeled polymer. For each filter, results were calculated based on the standard curve on that filter. The averages of one to five determinations (usually more than three) are shown.

^b AT/AT probes hybridize to bound AT sequences at approximately 10% the efficiency compared with the self-hybridization of most other sequences.

^c CG/CG probes hybridize to bound CG sequences at 1 to 3% the efficiency compared with the self-hybridization of most other sequences.

^d ND, Not detectable.

^e Single numbers in parentheses indicate number of determinations with the same value.

Drosophila species, e.g., chromosome 4 in *D. melanogaster* and chromosome 6 in *D. virilis*).

In addition to the conservation of distribution on the chromosomal level evident in the comparisons of different *Drosophila* species, the location of CA/TG sequences is highly conserved, at the level of polytene bands, within species, as shown by mapping in two *D. melanogaster* stocks. These stocks, gt-1 and gt-X11, have almost completely different sets of sites for a variety of mobile elements (11, 27, 38). In contrast, the maps of CA/TG sites in the two stocks were almost completely overlapping (29). Likewise, the locations of CT/AG and C/G sequences are conserved between the two stocks. On the other hand, maps of the three different sequence families overlap only occasionally (Fig. 4). Because of the packing of DNA in the chromosomes, even these few apparently coincident sites may not reflect contiguous repeats; however, some intersection of repeats has been demonstrated in the DNA sequence data base. One region of the X chromosome, between bands 13E and 17A, stands out because it contains very few sites for any of the repeats. The same region is underlabeled in *D. simulans*. We have not attempted to find a homologous region on the more distantly related species because of their divergent gene orders on the X chromosome.

Although it differs at the level of individual sites, the general chromosomal distribution of CT/AG (Fig. 2A) differs only slightly from that of CA/TG. CT/AG is enriched on the "ancestral" X chromosome in all species, but not to the

same extent as CA/TG. In *D. pseudoobscura* and *D. miranda*, CT/AG sequences are dramatically enriched on the "new" X^R arm, an arm that is autosomal in other *Drosophila* stocks. In *D. miranda*, the X₂ chromosome, which is only partially dosage compensating, shows distinct regions enriched in CT/AG. CA/TG shows a similar spotty pattern of hybridization on this chromosome; however, the patterns of hybridization of the two sequences overlap only occasionally.

CT/AG sequences are found only infrequently in the heterochromatin of the chromocenter; however, unlike CA/TG, they are not entirely excluded. In *D. melanogaster*, a circumscribed spot of hybridization is seen in the chromocenter. This may result from hybridization to CT/AG in the alpha-, rather than the beta-, heterochromatin. Alpha-heterochromatin contains repeated "satellite" DNAs which are underreplicated in polytene chromosomes (14) and detected as one or two small spots in the chromocenter. The other *Drosophila* species also show low levels of CT/AG in the chromocenter, but only in *D. melanogaster* are there suggestions that the repeat is in satellite DNA. CT/AG is underrepresented in the dot chromosome of every species except *D. virilis*.

The pattern of C/G sites is unmistakably different from that of CA/TG and CT/AG, although the features of the overall pattern still hold (Fig. 2B and C). In contrast to the many fine bands of hybridization seen on polytene chromosomes by CA/TG and CT/AG probes, C/G probes produce a

small number (5 to 20 per chromosome arm) of intensely labeled bands. There are some additional weaker bands, but even after very long exposures no C/G repeats can be detected in many chromosomal regions (Fig. 5B). Thus, C/G must occur in either relatively large blocks of closely linked segments or in segments of many small blocks that are not resolved by in situ hybridization. This is in sharp contrast to the CA/TG and CT/AG sequences which produce sharp bands of hybridization only after short autoradiographic exposures. After longer exposures, CA/TG and CT/AG hybrid is detected in most euchromatic chromosomal regions, although the enrichment on the X chromosome is still reflected by an increased level of hybridization (Fig. 5A).

Although in situ hybridization shows a limited number of C/G sites, hybridization of C/G to a Southern blot of *EcoRI*-cut DNA does not show hybridization to discrete fragments (data not shown). This suggests that each labeled band on the polytene chromosomes reflects C/G sequences interspersed with other sequences and clustered within some tens or hundreds of kilobases rather than a single long homopolymer stretch. However, these clusters may be in specialized regions of the genome; the failure to find numerous C/G repeats in the data base may result from a tendency to be excluded from regions near genes.

All species of *Drosophila* studied show a larger number of strongly labeled sites of C/G on the X chromosome than on the autosomes; however, there are a number of autosomal sites that are as heavy as individual sites on the X and there are also regions of the X that are entirely unlabeled. Even the chromosomes of *D. hydei*, which contains very little C/G (about the level of AT/AT sequences in some other species), show the typical pattern of C/G hybridization (Fig. 2C). After long exposures, *D. hydei* chromosomes show a small number of sites of C/G hybridization on the autosomes and a larger, but still limited, number of sites on the X chromosome.

In *D. pseudoobscura* and *D. miranda*, the chromosomes that have acquired dosage compensation more recently have an increased number of C/G loci. The new arm, X^R , of *D. pseudoobscura* is dramatically populated with regions containing C/G, while the ancestral X^L has only marginally more C/G loci than the autosomes. In *D. miranda*, the X_2 contains the highest number of C/G loci while the number of sites on the X^R and X^L resemble those in *D. pseudoobscura*. Once again, the set of C/G loci on the X_2 overlap but do not match the sets of loci containing either CA/TG or CT/AG.

FIG. 1. Occurrence of simple repeating sequences in a nucleic acid data base. The EMBL Nucleotide Sequence Data Library, Release 11.0, was searched for sequences of 10 base pairs or longer with no mismatches. Each repeating sequence and its complement were found, and the number in each size class was determined. The x axis shows length in base pairs [i.e., 10 means (dC-dA)₅ · (dT-dG)₅, etc.]. The y axis displays the numbers of members in each class. Numbers above a column show absolute values for classes containing more than 120 sequences; the y axis was set up to display clearly members of size classes larger than 20 base pairs. AT/AT sequences in mitochondrial genomes make up a large subpopulation and are indicated separately. The data base consists of 11,621 sequences and 12,189,783 nucleotides. Any sequence 10 bases long (and its complement) would be expected to be found 24 times in this data base if the nucleotide distribution were entirely random. Sequences 11 to 15 bases in length would be represented eight times. Sequences of more than 16 base pairs would be extremely improbable. Sequences from *Drosophila* spp. were not a large enough population to graph separately.

FIG. 2. Autoradiograph showing the chromosomal distribution of CT/AG and C/G. In each case the X chromosome (indicated) is the more heavily labeled. (A) Hybridization of [^3H]CT/AG to the chromosomes of *D. simulans* is characterized by general labeling on all chromosomes and by a striking concentration on the X. (B) Hybridization of [^3H]C/G to *D. simulans* chromosomes is also concentrated on the X but shows a more discrete pattern than the CT/AG hybridization. Individual sites on autosomes may be as heavily labeled as many X-chromosome sites; however, the number of sites on the X is strikingly enhanced. (C) Hybridization of [^3H]C/G to *D. hydei* chromosomes is lighter, as expected from the lower C/G content (Table 1). Only a very few regions are labeled after long autoradiographic exposures. The star marks a strongly labeled autosomal site. The chromocenter (C) shows little or no label. The lack of labeling of the chromocenter is also seen in species in which chromosome arms are more densely covered.

Although long C/G sequences are not common in the chromocenter, they are slightly more abundant than the other two repeats that have been discussed. Unlike CA/TG and CT/AG, C/G repeats can show some defined localization within the beta-heterochromatin; in *D. melanogaster* the bases of chromosomes 2R and 3L contain significant amounts of C/G sequences.

Repeats of AT/AT and A/T, although present in the genomes of *Drosophila* spp., are not detectable by in situ hybridization to polytene chromosomes. Even after very long exposures, no labeling is observed. A/T sequences are relatively rare, occurring an order of magnitude less than

other repeating sequences; however, they are observed at frequencies which might be expected to be detectable in situ in *D. virilis* and *D. hydei*, if not in other species, if they are located in polytenized regions of the chromosome arms. (Any sequences present in mitochondrial DNA or in the unpolytenized pericentric heterochromatin would be detected by Southern blot hybridization but not by in situ hybridization to polytene chromosomes.)

Regions that are becoming heterochromatic have not acquired additional mono- or dinucleotide repeats. The striking underrepresentation of mono- and dinucleotide repeats in the chromocenter suggests that this may be a general feature

FIG. 3. Schematic diagram of chromosome evolution of *Drosophila* species. Diagrams show the metaphase configurations of several *Drosophila* species. X chromosomes are shown in outline and Y chromosomes are shaded. The partially dosage compensating X_2 of *D. miranda* is indicated by arrows in both the male and female karyotypes and is shown paired with the neo-Y chromosome in the male karyotypes. The configuration of *D. virilis* is thought to be that of the ancestral species and is the same as that of *D. hydei* (not shown), although the two are not closely related. An estimate of the time since these species have diverged is shown below the karyotypes. my, Millions of years

FIG. 4. Autoradiogram comparing the hybridization of CA/TG, C/G, and CT/AG to the X chromosome of *D. melanogaster*. Chromosomes are from the gt-1 stock. Autoradiographic exposures were chosen to give roughly equivalent total labeling for the different polymers. Region 13-17, which is underlabeled with each probe, is indicated. Equivalent regions on the X chromosome are connected. Although the overall pattern of hybridization is similar for each probe, specific features are quite different. (A) CA/TG hybridization to an X chromosome. (B) C/G hybridization to an X chromosome and autosome 2R. The striking concentration of label over the distal end of the X is obvious, and there are a number of proximal sites. The X chromosome is broken at the "weak spot" in 11A; it is interesting to note a fan of autoradiographic grains extending from both sides of the break, suggesting that this unusual structure is rich in C/G repeats. (C) CT/AG hybridization to an X chromosome. Part of autosome 3R is also seen. With this probe, the concentration of repeats on the X chromosome is less marked, but still detectable. The scatter of grains surrounding the labeled regions of the chromosomes is characteristic of this probe.

FIG. 5. Comparison of long autoradiographic exposures of hybridizations with CT/AG and C/G to the tip of the X^R chromosome of *D. pseudoobscura*. The two repeats show distinctly different patterns of hybridization. (A) Although at lower exposures, CT/AG appears to localize in a large number of fine bands; after long exposures, an almost uniformly saturated labeling of the X^R tip is seen. (B) C/G sequences appear localized in six major and a number of minor sites on the tip of X^R , with no evidence of other lesser sites in the intervening regions, even after our longest exposures. A piece of an autosome is also shown. The autosomes have fewer C/G sequences but the sequences show the same well-defined pattern seen on the X chromosome.

of heterochromatin. The *D. miranda* chromosome set provides an opportunity to study ancestral autosomes that have evolved dosage compensation. This species also offers a chance to study a chromosome that is becoming heterochromatic; an ancestral autosome has become fused to the Y chromosome, forming the neo-Y. The homolog of the ancestral autosome is now called the X_2 because it segregates as an X chromosome. The female has two X_2 chromosomes; in the male the X_2 is paired with the neo-Y and is effectively hemizygous for genes that are inactivated by the heterochromatization of the neo-Y. Hence, we can compare former homologs, one acquiring heterochromatic regions and the other acquiring regions that dosage compensate (10, 32, 33). The heterochromatic nature of the neo-Y is detected in metaphase chromosomes by darker staining. In polytene chromosomes, the ancestral Y does not replicate, while the newly acquired arm polytenizes into a tangled mass resembling beta-heterochromatin.

The heterochromatization of the neo-Y arm has been accompanied by the acquisition of many repeated sequences (32), but these sequences, like other heterochromatin, do not include mono- and dinucleotide repeats. CA/GT and CT/AG are represented at levels approximately equal to or slightly less than the levels on the euchromatic autosomal arms (Fig. 6). C/G is almost entirely absent from the neo-Y, which shows none of the intensely labeled areas seen on the other chromosomes. The tangling of the neo-Y prevents clear morphological dissection but it appears that, at least for CT/AG, some of the sites on the X_2 are found on the neo-Y, where they show reduced hybridization. Other X_2 sites are not detected on the neo-Y. The shared sites may represent loci that have been conserved from a common ancestor and have become amplified in the X_2 .

DISCUSSION

The conserved chromosomal locations of simple repeating DNAs make it attractive to postulate roles for these sequences in chromosome structure and function. Three of the six possible mono- and dinucleotide repeats are found as

long, relatively uninterrupted runs distributed nonrandomly throughout the euchromatin of several *Drosophila* species. The moderate hybridization stringencies used for both in situ and filter hybridizations require approximately 20 to 30 nucleotides of homology to form a stable hybrid. In addition, the autoradiographic exposures for in situ hybridization are chosen, on the basis of other work (M. L. Pardue, unpublished data), to require some 50 to 100 base pairs of hybrid (per chromatid) in a given region to form a detectable band on the polytene chromosome. A moderate degree of mismatch interruption is possible, especially in long repeats; however, absolutely no cross-hybridization between the different probes can be detected. Thus, the long repeats detected in the filter hybrids are probably at least 20 to 30 nucleotides, while the hybrids detected on the polytene chromosomes probably require several times that number in close proximity.

In contrast to CA/TG, CT/AG, and C/G, repeats A/T are an order of magnitude less abundant in the genome and are not detected in the in situ hybridizations we performed. A/T sequences may be concentrated in unpolytenized alpha heterochromatin with satellite DNA. Two of the simple repeats, AT/AT and CG/CG, are poorly detected by hybridization, probably because they have the ability to form intramolecular hairpin structures. In spite of this, a significant amount of AT/AT can be detected by hybridization to purified DNA. No AT/AT sequences were detected by in situ hybridization, suggesting that stretches of this copolymer are either shorter or less clustered than CA/TG or CT/AG repeats. In addition, many of the AT/AT repeats found in the data base are from mitochondrial DNA. Mitochondrial DNA from all species of *Drosophila* has a large A+T-rich segment (13, 28) which has AT/AT repeats in the one species, *D. yakuba*, in which it has been sequenced (9). Mitochondrial DNA sequences would be detected in the total DNA used for the filter hybridizations but would not be found in the in situ hybridization to polytene chromosomes. CG/CG repeats were not detected by either hybridization technique. The absence of CG/CG may be due to the poor hybridization efficiency; however, other evidence supports the idea that the genome actually does not contain long CG/CG repeats (16, 26; see below).

Another source of information about these mono- and dinucleotide repeats is the DNA sequence data base. This is strongly biased toward coding regions and the sequences flanking them. It remains to be seen how representative this is of the rest of the genome. However, an analysis of sequences now in the data base yields numbers which are generally in accord with our results. CA/TG and CT/AG sequences of up to 100 base pairs can be found, and uninterrupted stretches of >30 base pairs are common. The representation of AT/AT and A/T sequences falls off dramatically, and stretches of more than 20 base pairs are relatively rare. CG/CG repeats are rare and very short.

There is a noticeable disagreement between the representation of C/G in the data base and our hybridization results; C/G sequences are found only rarely in the data base, while we find them abundantly in *Drosophila* genomes. The difference may be due to a cloning artifact (repeating sequences are markedly unstable in plasmids). Alternatively, this contradiction may be the result of the clustering and sequestering of long C/G repeats, observed in the polytene chromosomes. The genomic regions containing these clusters may not yet have attracted the interest of sequencers. It is also evident that A/T, which is a relatively rare sequence in *Drosophila* spp., is very prevalent in the data base. Filter

FIG. 6. Comparison of the X_2 and the neo-Y chromosomes of *D. miranda*. Although derived from the same autosome, these chromosomes are quite diverged. The X_2 has acquired ability to partially dosage compensate and a higher level of mono- and dinucleotide repeats than seen on autosomes. The neo-Y is now partially heterochromatic and has acquired a higher level of repeated sequences but the acquired sequences do not include mono- and dinucleotide repeats, which are present on the neo-Y at levels below or equal to levels on the autosomes. One manifestation of the heterochromatization of the neo-Y is its amorphous structure in polytene nuclei. This makes it impossible to directly compare regions with its former homolog X_2 . X, Tip of X_2 ; Y, neo-Y. (A) CA/TG sequences are prominent of the X_2 and present at a much lower level on the neo-Y. (B) C/G sequences are identified at a number of locations on the X_2 but, even after a very long autoradiographic exposure, the number of grains over the neo-Y is only slightly above background. (C) CT/AG sequences hybridize strongly to the X_2 and less strongly to the neo-Y. AT/AT and A/T sequences have also been tested and show no hybridization with the neo-Y.

hybridization shows that the concentration of A/T is two orders of magnitude higher in yeast DNA than in *D. melanogaster* DNA (K. Lowenhaupt, unpublished data). AT/AT repeats may also be more prevalent in genera other than *Drosophila*.

Some sequences are not fully replicated in polytene chromosomes and are therefore not detected by in situ hybridizations. These sequences include the Y chromosome and the satellite DNAs of alpha-heterochromatin (14). Our results suggest that the alpha-heterochromatin of *D. melanogaster* may also contain some long A/T and CT/AG repeats. When DNA has been cleaved by restriction enzymes, both

of these polymers hybridize strongly to very large fragments, suggesting many repeats clustered without intervening restriction sites, typical of satellite DNA (data not shown). A/T is not detected on polytene chromosomes, while CT/AG probes show one or two small dots of hybridization within the chromocenter, resembling the patterns of known satellites in alpha-heterochromatin.

The three sequence families detected in polytene chromosomes, CA/TG, CT/AG, and C/G, are all distributed over the euchromatic arms and enriched on the X chromosome and on other chromosome arms that have evolved the ability to dosage compensate. They are also markedly underrepre-

sented in the beta-heterochromatin, although the chromocenter is not as entirely free of CT/AG and C/G as it is of CA/TG. The underrepresentation of mono- and dinucleotide repeats in beta-heterochromatin is extended to the sequences involved in the heterochromatization of the new arm of the Y in *D. miranda*. Both the beta-heterochromatin and the new arm of the Y are enriched for repeated sequences (32, 37), yet these repeated sequences do not include homopolymer or alternating copolymer repeats.

The increased level of repeat sequences that lack homopolymer and alternating copolymer repeats on the neo-Y contrasts with its former homolog, now called the X₂. The X₂ has acquired the ability to dosage compensate in some regions (10, 33). It has also acquired increased levels of CA/TG, CT/AG, and C/G. The homolog that is now part of the neo-Y has lost much of its organized polytene structure; the X₂ has not undergone any gross change in chromatin structure.

Mapping of individual sites of CA/TG, CT/AG, and C/G on *D. melanogaster* chromosomes shows that sites for each of the three repeats are conserved between distantly related stocks. We conclude that these sequences are not being stochastically amplified and deleted, except, perhaps, on an evolutionary time scale. The relative stability of repeat sites is also supported by the recent discovery that CA/TG sequences have conserved locations bracketing ribosomal genes in various mammalian species (5). Chromosomal evolution in *Drosophila* spp. shows rearrangements within chromosomes but rarely rearrangements between chromosomes (30). Homology of chromosomes between species is well established, but little comparative mapping of specific sites has been done. For that reason we have not attempted to compare sites of repeats in different species, except for comparisons between the sibling species, *D. melanogaster* and *D. simulans*, which have homosequential chromosomes.

The most dramatic observation about the general pattern of distribution of CA/TG is an increase on chromosome arms that have acquired the ability to dosage compensate (29). As discussed in that study, this is evidence for a general difference in the chromatin in regions capable of dosage compensation which may be part of the mechanism of dosage compensation. A similar correlation exists between dosage compensation and increases in CT/AG and C/G repeats. The three sequence families do not map to the same locations. It is therefore probable that all of these families (and possibly other repeats not identified in this study) share a role. Since each of the sequences reported here is capable of at least two strikingly different interconvertible DNA structures, it is attractive to propose that the role is to confer structural lability on the chromatin structure.

The role which a topologically labile sequence, CA/TG, could play in dosage compensation has been discussed previously (29). Both CT/AG and C/G readily adopt a non-B conformation when subjected to negative supercoiling. This conformation, first detected as sensitivity to the single-strand nuclease S1 (21), has been extensively studied, and several models have been proposed (7, 19, 23, 31). The general similarities seen in the chromosomal distributions of CA/TG, CT/AG, and C/G may be the result of an important attribute of these sequences, their common, dramatic structural responsiveness to negative superhelical density.

CA/TG, CT/AG, and C/G repeats are abundant and generally conserved in the genome. A/T and AT/AT repeats are shorter and more sparsely dispersed, while CG/CG repeats of any length appear to be essentially absent. It may be that CG/CG forms Z-DNA too readily and is therefore selected

against. A+T-rich regions melt easily; it is possible that, in a system in which single-stranded DNA is used for other things, A+T-rich DNA cannot be exploited for topological lability. Because at least some repeating mono- and dinucleotide are extremely recombinogenic (6, 36), it is likely that longer stretches are continually trimmed and excised unless there is pressure to maintain them. Such shortening is seen in bacteria, in which runs above a certain size (which varies from sequence to sequence) are very unstable.

Evidence that repeating mono- and dinucleotides can be recombinogenic adds another point of interest to their chromosomal distribution. The regions where we find few, if any, of these repeats are also known not to undergo meiotic recombination. These regions include pericentric heterochromatin on all chromosomes as well as the dot chromosome in most species. It is notable that the dot chromosome in *D. virilis* (chromosome 6), in which we detect CA/TG and CT/AG, is also the only dot chromosome in which recombination is detected (8). It would be particularly interesting to know whether recombination of the X chromosome is more frequent than recombination on autosomes of the same species, as might be predicted from comparison of the amounts of simple repeats. Unfortunately, comparisons of recombination rates between different chromosomes can be made only if actual DNA lengths are known. The regions for which appropriate DNA measurements have been made are not yet extensive enough to provide a representative sample.

The differences between species in the relative amounts of mono- and dinucleotide repeats provide a possible explanation for a question that has puzzled geneticists. X chromosomes in different *Drosophila* species contain the same genes, although the gene order is not constant. Because of the genetic equivalence, X chromosomes are assumed to contain roughly equivalent amounts of DNA, yet the genetic length of the X chromosome from the tip to the base (a reflection of the relative amount of recombination) differs from species to species. The length is 66 crossover units in *D. melanogaster* (22), 116 in *D. hydei* (17), 170 in *D. virilis* (2). In *D. pseudoobscura*, the original arm of the X is 73.1 map units (34). Our hybridization data show a positive correlation between the relative level of simple repeats in the genome of each species (Table 1) and the genetic length of the X chromosome.

The recombinogenicity of mono- and dinucleotide repeats may be of importance either as a second independent function or as part of the mode by which the repeats were imported to the X^R and X₂. It is possible that the lack of meiotic recombination in heterochromatic regions is due to a lack of the recombinogenic repeat sequences. On the other hand, these sequences may have become dispersed through the genome by a recombination-based mechanism and thus did not enter the heterochromatin or the dot chromosome because the regions do not participate in the recombination.

The X^R chromosomes of *D. pseudoobscura* and *D. miranda* and the X₂ of *D. miranda* contain different relative amounts of CA/TG, CT/AG, and C/G than do the ancestral X chromosomes (the X^L of *D. miranda* and *D. pseudoobscura* and the entire X of the other species). The "recent" X chromosome arms have considerably more CT/AG and C/G than the ancestral Xs and relatively less CA/TG. If, as we suggest, a sufficient amount of any of these sequences, appropriately distributed, would suffice, then differences in relative amounts may reflect the relative ease with which each sequence can be imported.

ACKNOWLEDGMENTS

K.L. has been a Fellow of the Arthritis Foundation. This work was supported by grants from the American Cancer Society (to M.L.P. and to A.R.).

Data base analyses were done through the Whitaker College Computing Facility. We thank Daniel Sullivan and Betagen Co. for assistance and the use of the Betagen Blot analyzer.

LITERATURE CITED

1. Abraham, I., and J. C. Lucchesi. 1974. Dosage compensation of genes on the left and right arms of the X chromosome of *Drosophila pseudoobscura* and *Drosophila willistoni*. *Genetics* 78:1119-1126.
2. Alexander, M. L. 1976. The genetics of *Drosophila virilis*, p. 1365-1427. In M. Ashburner and E. Novitski (ed.), *The genetics and biology of Drosophila*, vol. 1c. Academic Press, Inc., New York.
3. Aronson, J. F., G. T. Rudkin, and J. Schulz. 1954. A comparison of the giant X chromosomes in male and female *Drosophila melanogaster* by cytophotography in the ultraviolet. *J. Histochem. Cytochem.* 2:458-459.
4. Baker, R. F. 1977. Binding of DNA to cellulose nitrate filters under denaturing conditions. *Anal. Biochem.* 78:569-571.
5. Braaten, D. C., J. R. Thomas, R. D. Little, K. R. Dickson, I. Goldberg, D. Schlessinger, A. Ciccodicola, and M. D'Urso. 1980. Locations and contexts of sequences that hybridize to poly(dG-dT) · (dC-dA) in mammalian ribosomal DNAs and two X-linked genes. *Nucleic Acids Res.* 16:865-881.
6. Bullock, P., J. Miller, and M. Botchan. 1986. Effects of poly[d(pGpT) · d(pApC)] and poly[d(pCpG) · d(pGpG)] repeats on homologous recombination in somatic cells. *Mol. Cell. Biol.* 6:3948-3953.
7. Cantor, C. R., and A. Efstradiatis. 1984. Possible structures of homopurine-homopyrimidine S1-hypersensitive sites. *Nucleic Acids Res.* 12:8059-8072.
8. Chino, M., and H. Kikkawa. 1933. Mutants and crossing over in the dot-like chromosome of *Drosophila virilis*. *Genetics* 18:111-116.
9. Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* 22:252-271.
10. Das, M., D. Mutsuddi, A. K. Duttgupta, and A. S. Mukherjee. 1982. Segmental heterogeneity in replication and transcription of the X₂ chromosome of *Drosophila miranda* and conservativeness in the evolution of dosage compensation. *Chromosoma* 87:373-388.
11. Dawid, I. B., E. O. Long, P. P. DiNocera, and M. L. Pardue. 1981. Ribosomal insertion-like elements in *Drosophila melanogaster* are interspersed with mobile elements. *Cell* 25:399-408.
12. Dowsett, A. P., and M. W. Young. 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 79:4570-4574.
13. Fauron, C. M.-R., and D. R. Wolstenholme. 1980. Interspecific diversity of nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules of *Drosophila mauritiana*, *Drosophila melanogaster* and *Drosophila simulans*. *Nucleic Acids Res.* 8:5391-5410.
14. Gall, J. G., E. H. Cohen, and M. L. Polan. 1971. Repetitive DNA sequences in *Drosophila*. *Chromosoma* 33:319-344.
15. Gross, D., and W. T. Garrard. 1986. The ubiquitous potential Z-forming sequence of eucaryotes, (dT-dG)_n · (dC-dA)_n, is not detectable in the genomes of eubacteria, archaebacteria, or mitochondria. *Mol. Cell. Biol.* 6:3010-3013.
16. Hamada, H., M. G. Petrino, and T. Kakunaga. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* 79:6465-6469.
17. Hess, O. 1976. Genetics of *Drosophila hydei* Sturtevant, p. 1343-1363. In M. Ashburner and E. Novitski (ed.), *The genetics and biology of Drosophila*, vol. 1c. Academic Press, Inc., New York.
18. Huijser, P., W. Hennig, and R. Dijkhof. 1987. Poly (dC-dA/dG-dT) repeats in the *Drosophila* genome: a key function for dosage compensation and position effects? *Chromosoma* 95:209-215.
19. Johnson, D., and A. R. Morgan. 1978. Unique structures formed by purimidine-purine DNAs which may be four-stranded. *Proc. Natl. Acad. Sci. USA* 75:1637-1641.
20. Koo, H.-F., H.-M. Wu, and D. M. Crothers. 1986. DNA bending at adenine · thymine tracts. *Nature (London)* 320:501-506.
21. Larsen, A., and H. Weintraub. 1982. An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chicken globin chromatin. *Cell* 29:609-622.
22. Lindsley, D. L., and E. H. Grell. 1972. Genetic variations of *Drosophila melanogaster*. Carnegie Inst. Washington Publ. 627, p. 443.
23. Lyamichev, V. I., S. M. Mirkin, and M. D. Frank-Kamenetskii. 1985. Structure of (dG)_n · (dC)_n under superhelical stress and acid pH. *Biol. Struct.* 3:327-338.
24. McKnight, R. H. 1939. The sex determining mechanism of *Drosophila miranda*. *Genetics* 24:180-201.
25. Mukherjee, A. S., and S. N. Chatterjee. 1975. Chromosomal basis of dosage compensation in *Drosophila*. VIII. Faster replication and hyperactivity of both arms of the X-chromosome in males of *Drosophila pseudoobscura* and their possible significance. *Chromosoma* 53:91-105.
26. Nordheim, A., M. L. Pardue, L. M. Weiner, K. Lowenhaupt, P. Scholten, A. Moller, A. Rich, and B. D. Stollar. 1986. Analysis of Z-DNA in fixed polytene chromosomes with monoclonal antibodies that show base sequence-dependent selectivity in reactions with supercoiled plasmids and polynucleotides. *J. Biol. Chem.* 261:468-476.
27. Pardue, M. L., and I. B. Dawid. 1981. Chromosomal locations of two DNA segments that flank ribosomal insertion-like sequences in *Drosophila*: flanking sequences are mobile elements. *Chromosoma* 83:29-43.
28. Pardue, M. L., J. M. Fostel, and T. R. Cech. 1984. DNA-protein interactions in the *Drosophila virilis* mitochondrial chromosome. *Nucleic Acids Res.* 12:1991-1999.
29. Pardue, M. L., K. Lowenhaupt, A. Rich, and A. Nordheim. 1987. (dC-dA)_n · (dG-dT)_n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO J.* 6:1781-1789.
30. Patterson, J. T., and W. S. Stone. 1952. Evolution in the genus *Drosophila*. Macmillan Publishing Co., Inc., New York.
31. Pulleyblank, D. E., D. B. Haniford, and A. R. Morgan. 1985. A structural basis for S1 nuclease sensitivity of double-stranded DNA. *Cell* 42:271-279.
32. Steinemann, M. 1982. Multiple sex chromosomes in *Drosophila miranda*: a system to study the degeneration of a chromosome. *Chromosoma* 86:59-76.
33. Strobel, E., C. Pelling, and N. Arnheim. 1978. Incomplete dosage compensation in an evolving *Drosophila* sex chromosome. *Proc. Natl. Acad. Sci. USA* 75:931-935.
34. Sturtevant, A. H., and C. C. Tan. 1937. The comparative genetics of *Drosophila pseudoobscura* and *D. melanogaster*. *J. Genet.* 34:415-432.
35. Sullivan, D. E., P. E. Auron, G. J. Quigley, P. Watkins, J. E. Stanchfield, and C. Bolon. 1987. The nucleic acid blot analyzer. I. High speed imaging and quantitation of ³²P-labeled blots. *Biotechniques* 5:672-678.
36. Treco, D., and N. Arnheim. 1986. The evolutionarily conserved repetitive sequence d(TG · AC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* 6:3934-3947.
37. Young, B. S., A. Pession, K. L. Traverse, C. French, and M. L. Pardue. 1983. Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell* 34:85-94.
38. Young, M. W., and H. E. Schwartz. 1981. Nomadic gene families in *Drosophila*. Cold Spring Harbor Symp. Quant. Biol. 45:629-640.