J. Mol. Biol. (1982) **158**, 17–35

Dynamics of Concerted Evolution of Ribosomal DNA and Histone Gene Families in the *melanogaster* Species Subgroup of *Drosophila*

ENRICO COEN, TOM STRACHAN AND GABRIEL DOVER

Department of Genetics, University of Cambridge Cambridge CB2 3EH, England

(Received 18 December 1981)

Many multigene and non-coding families of DNA reveal unexpected high levels of sequence homogeneity (concerted evolution). In order to assess the rates and mechanisms of concerted evolution, we have analysed sequence variation in functionally diverse regions of the ribosomal DNA and histone gene families in seven sibling species of the *Drosophila melanogaster* species subgroup. Variant sequences of the non-transcribed and transcribed rDNA spacers have become fixed in both the X and Y chromosome rDNA clusters in a diagnostic manner for each species. Similarly, species-specific variants have become fixed in the spacer and coding regions of the histone cluster. Variation in non-transcribed spacer length, within and between species, is based on multiples of the periodicity of an internal region of repetition within the non-transcribed spacer. These observations, in conjunction with the genetic analysis of the extensive polymorphism in non-transcribed spacer length and copy-number in *D. melanogaster* (Coen *et al.*, 1982) suggest that unequal exchange is a relatively rapid process and results in a turnover of sequences at several periodicities within the arrays. The rate of turnover can contribute substantially to the fixation of variants within an array and the stochastic process of drift can contribute to the fixation of an array in a population. However, the observation that regions of rDNA and histone repeats have become homogeneous for variants that can discriminate between the two most closely related and recently diverged species suggests that additional processes might be required to drive variants to complete fixation. Data on the distribution of the type I and type II introns of the rDNA implies a process of continual re-insertion into newly fixed rDNA variants during the homogenization process. Comparative data on the forces responsible for the evolutionary progression of variants in families of larger copy-number and wider karyotypic distribution in the same group of species are described in the accompanying paper (Strachan *et al.*, 1982).

1. Introduction

A feature of many multigene and non-coding families is that they are subject to a continuous process of homogenization that imparts a greater within-species homogeneity than between-species homogeneity. This process, known as concerted evolution, has been found in multigene families of 28 S and 18 S ribosomal DNA, 5 S rDNA, globins, immunoglobulins, heat-shock genes, histones and repeated

0022–2836/82/170017–19 $03.00/0

sequences of unknown function (Brown *et al.*, 1972; Arnheim *et al.*, 1980; Dover & Coen, 1981; Ford & Brown, 1976; Zimmer *et al.*, 1980; Lauer *et al.*, 1980; Baltimore, 1981; Leigh Brown & Ish-Horowicz, 1981; Kedes, 1979; Brown & Dover, 1981; Pan *et al.*, 1981; Moore *et al.*, 1978; Flavell, 1982). For reviews, see Jeffreys (1982); Long & Dawid (1980); Fedoroff (1978); and Dover *et al.* (1982).

In order to understand the dynamics of concerted evolution, we have chosen to study genic and non-genic repetitive families of seven sibling species of the *Drosophila melanogaster* species subgroup. These species are of particular interest, since their genetic and phylogenetic relationships are becoming well-defined (see Dover *et al.*, 1982 for references), and the relative rates of evolution of several biological components can be assessed.

In this paper, we describe studies on the rates of concerted evolution within the ribosomal and histone gene clusters that complement studies on abundantly repeated tandem sequences in the accompanying paper and studies on rDNA polymorphism within *D. melanogaster* (Coen *et al.*, 1982). A fine-structural analysis has been made of the relative rates of homogenization between functionally different regions within the compound histone and rDNA units. In the case of rDNA, for example, rate comparisons are feasible between the genes, the introns, the transcribed and non-transcribed spacers and the internal region of repetition within the latter, for both the X and Y chromosome clusters (see Fig. 1).

2. Materials and Methods

(a) *Construction of clones*

The rDNA clone pDm238 is an 11·5 kb† *Eco*RI fragment cloned into the plasmid pBR322 and was a kind gift from D. Glover (Roiha *et al.*, 1981). The spacer clone pDm103HH2 was constructed by subcloning a large *Hin*dIII/*Hae*III fragment from cDm103 (Glover & Hogness, 1977) into the plasmid pAT153. The type I intron clone pC225 contains 5 kb of the type I insertion, cloned into pBR322 and was constructed by S. Kidd (Roiha & Glover, 1980). The type II intron clone pCDm207CD consists of 2 *Eco*RI fragments from the type II insertion of cDm207, cloned into pBR322 (Roiha & Glover, 1980). The histone clone, cDm500 (Lifton *et al.*, 1977) was kindly supplied by Bob Karp.

(b) *DNA extraction and purification*

DNA was prepared from an equal mixture of male and female flies as described by Barnes *et al.* (1978). In those cases where only 8 flies were used for the extraction, a modification of a method of Ish-Horowicz was followed (Coen *et al.*, 1982), except that all extraction volumes were doubled.

(c) *Enzyme reactions*

Restriction enzymes were obtained from Bethesda Research Labs, Boehringer–Manheim Biochemicals or New England Nuclear. *Ava*I was a kind gift from G. Roizes. Digestions were carried out as recommended by the supplier. Internal standards of bacteriophages λ and M13 DNA were used to check that digestions were complete.

Nick translations were carried out essentially according to Rigby *et al.* (1977). Labelling was with [^{32}P]dATP (400 Ci/mmol; Amersham) to a specific activity of 10^7 to 10^8 cts/min per μg.

† Abbreviations used: kb, 10^3 base-pairs as appropriate; bp, base-pair; NTS, non-transcribed spacer; ITS, internal transcribed spacer.

(d) *Electrophoresis*

Agarose gels were run in Tris-borate (89 mM-Tris, 2·5 mM-Na_2EDTA, 89 mM-boric acid) overnight. Gels were blotted into Schleicher and Schull BA85 nitrocellulose filters according to the method of Wahl *et al.* (1979). Restriction fragments were recovered from preparative gels essentially according to the hydroxyapatite adsorption method of Tabak & Flavell (1978).

(e) *Filter hybridization*

Following transfer of DNA to nitrocellulose filters, the filters were blotted dry and baked at 80°C for 2 h. Radioactively labelled probe (5×10^6 to 10×10^6 cts/min) was boiled for 10 min in 15 ml of a hybridization solution containing 5 × SSC (SSC is 0·015 M-NaCl, 0·15 M-sodium citrate), 0·5% (w/v) sodium dodecyl sulphate and 50% (v/v) formamide. The solution was then added to a dried filter and hybridization was conducted at 42°C overnight. Filters were washed extensively in 3 mM-Tris (unneutralized), dried and autoradiographed.

Filters were reprobed after allowing sufficient time for their residual radioactivity to decay. The filters were washed for 15 to 30 min in 30 mM-NaOH, neutralized in 3 mM-Tris, pre-incubated for 1 h in 20 ml hybridization solution and challenged with a new radioactivity labelled probe.

3. Results

(a) *Ribosomal DNA non-transcribed spacer length variation*

In *D. melanogaster* there are approximately 200 tandem rDNA repeats in each nucleous organizer on the X and Y chromosomes, organized as shown in Figure 1 (see the legend to Fig. 1; for a review, see Long & David, 1980).

Digestion of *D. melanogaster* rDNA with *Hae*III produces a fragment containing the complete NTS together with part of the external transcribed spacer and part of the 28 S gene (Fig. 1). *Hae*III-digested total DNA of eight females from each of the seven sibling species was probed with a NTS clone pDm103HH2 (Fig. 2(a)). The pattern of hybridizations obtained, together with those obtained from restrictions with *Alu*I (see later) and *Hin*dIII (not shown) indicate that the differences in

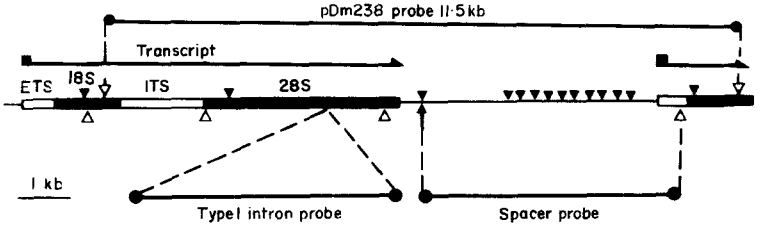


FIG. 1. Map of a repeating unit of ribosomal genes in *D. melanogaster* (for full references, see Long & Dawid, 1980). Regions that give rise to the 18 S and 28 S RNAs are shown as filled bars; transcribed spacers are distinguished as external (ETS) or internal (ITS). The length and direction of the primary transcript is shown by an arrow above the coding region. The 28 S genes of *D. melanogaster* may have 1 of 2 non-homologous introns (type I and type II). In Oregon R stocks there are type II introns in about 15% of the units of the X and Y chromosomes, and type I introns in about 50% of the X chromosome units. Clones used as probes are shown above and below the restriction map and consist of: a total rDNA repeat lacking an intron, pDm238; a type I intron, pC225; and a *Hin*dIII/*Hae*III spacer fragment, pDm103HH2. The restriction sites are: ⏚, *Eco*RI; △, *Hae*III; ↑, *Hin*dIII; ▼, *Alu*I. The position of only those sites in, or flanking the NTS or ITS are shown for *Hae*III and *Alu*I. Only 1 *Hin*dIII site is shown, to demonstrate the origin of the spacer clone pDm103HH2.

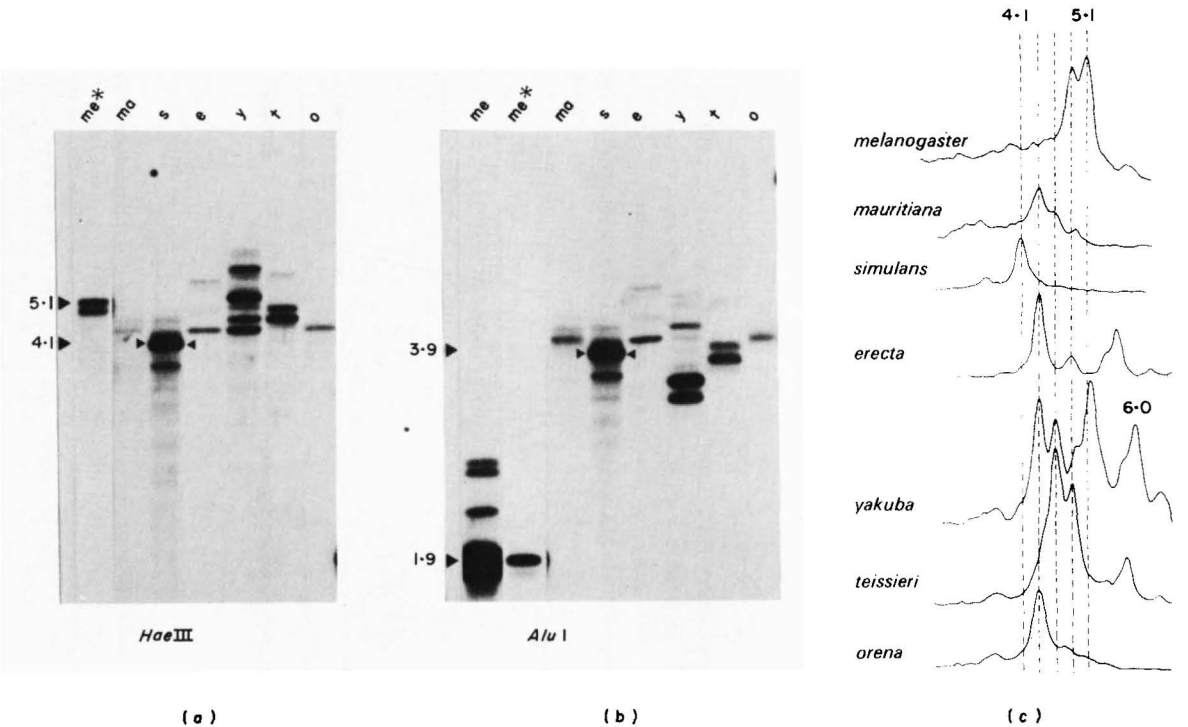


FIG. 2. (a) Autoradiogram of total female DNA from the 7 sibling species after digestion with *Hae*III, fractionation on 0·8% agarose gel and hybridization to a ^{32}P-labelled NTS clone, pDm103HH2. Each track contains the DNA from 1/3 of an 8-fly extraction. The abbreviations used for each species (here and in the following Figures) are: me, *D. melanogaster*; ma, *D. mauritiana*; s, *D. simulans*; e, *D. erecta*; y, *D. yakuba*; t, *D. teissieri*; o, *D. orena*. Arrows on the left indicate fragment sizes in kb. An asterisk indicates an 8-fold shorter time of exposure for the autoradiogram. Differences between species in the overall intensity of hybridization may be due to the different amounts of DNA loaded onto the gel, which depends on the total amount of DNA extracted per fly, differences in sequence homology and the differing amounts of rDNA per haploid genome (Tartof, 1979). (b) Autoradiogram of total female DNA from the 7 sibling species after digestion with *Alu*I, fractionation on 0·8% agarose gel and hybridization to a ^{32}P-labelled NTS clone, pDm103HH2. Each track contains the DNA from 1/3 of the same 8-fly extraction as used in (a). (c) Microdensitometer tracing of the autoradiogram shown in (a). Broken lines are separated by 250 bp intervals and the numbers indicate sizes (in kb).

fragment lengths within and between species are due to changes in the length of the region separating the same two flanking *Hae*III sites. Quantitative microdensitometry of Figure 2(a) reveals three significant features (Fig. 2(c)). First, fragment lengths both within and between species often differ by 250 base-pairs or integrals of 250 bp, suggesting that fragment length variation is due to changes in the number of the 250 bp repeats within the NTS rather than to changes in the length of the 28 S or external transcribed spacer components of the *Hae*III fragment. Secondly, all the species have one or more major fragments in the range 4·1 to 5·1 kb, with a few additional fragments outside this range in some species. Thirdly, several of the species show very little overlap in their fragment sizes (e.g. between *Drosophila mauritiana* and *Drosophila simulans*), indicating that different NTS lengths have replaced ancestral arrays in these species.

(b) *Concerted evolution of non-transcribed spacer sequences*

The NTS of *D. melanogaster* rDNA is sensitive to *Alu*I, which recognizes about 7 to 12 sites at 250 bp intervals (Long & Dawid, 1979; see Fig. 1). *Alu*I-restricted total DNA, from an equal mixture of males and females from each of the seven species, was hybridized with the NTS probe pDm103HH2 (Fig. 3(a)). In addition to the expected 250 bp band, the rDNA of *D. melanogaster* shows two major bands, at 1·9 kb and 1·4 kb, corresponding to the left and right of the *Alu*I region, respectively. None of the other species shows any band at 250 bp and the majority of their NTSs are left as large *Alu*I-resistant fragments. In order to resolve these large fragments, *Alu*I-digested female DNA of each species was fractionated on a

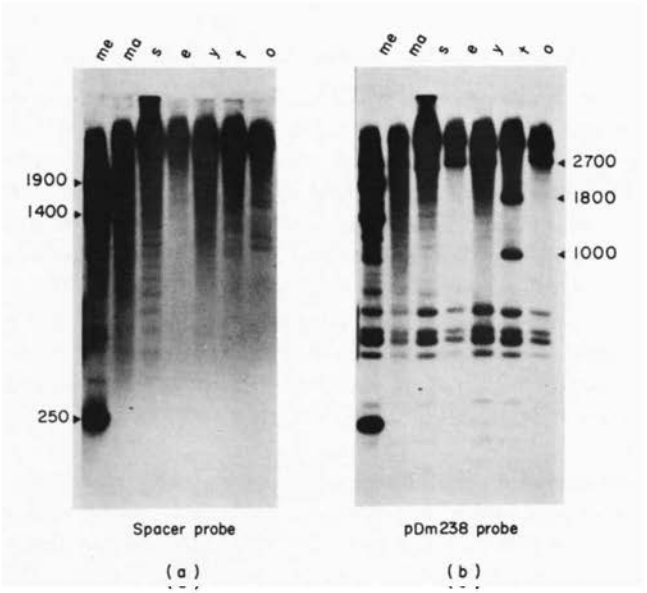


FIG. 3. (a) Autoradiogram of total male/female mixed DNA from the 7 species after digestion with *Alu*I, fractionation on a 2% agarose gel and hybridization with a ^{32}P-labelled NTS clone, pDm103HH2. Sizes are in base-pairs. (b) Autoradiogram of the same filter as shown in (a) after removal of the spacer probe, and reprobing with a cloned total rDNA repeat, pDm238. Sizes are in base-pairs.

lower percentage agarose gel (Fig. 2(b)). Comparison of Figure 2(a) and (b) shows that the *Hae*III and *Alu*I fragments differ by a constant factor of 200 bp for each of the *Drosophila* species *D. mauritiana*, *D. simulans*, *D. erecta* and *D. orena*, confirming that the NTSs of these species are refractory to *Alu*I. However, the *Alu*I and *Hae*III patterns of hybridization in each of the *Drosophila* species *D. teisseiri* and *D. yakuba* do not differ by a constant 200 bp factor, due to the presence of additional *Alu*I sites in the regions flanking the internal repetition of the NTSs. In marked contrast to *D. melanogaster*, the above results indicate that none of the six species contains *Alu*I sites at 250 bp intervals. Microdensitometric analysis of the *Alu*I digests shows that the internal 250 bp repeats within all the NTSs of *D. melanogaster* contain *Alu*I sites.

Probing of *Eco*RI-digested total female DNA of the species with clone pDm238 (Fig. 1) shows that five of the species have major bands in the range 11 to 17 kb, which corresponds to the length expected of total rDNA repeats (Fig. 4(a)). However, *D. mauritiana* and *D. simulans* have *Eco*RI sites within their NTSs (Fig. 4(c)). In addition to the sites shown in Figure 4(c), some of the NTSs of *D. mauritiana* contain a number of other *Eco*RI sites at intervals of 100 bp (Fig. 4(b)). A similar 100 bp periodicity is seen in the *Alu*I digests of *D. simulans* DNA (Fig. 3(a)), which suggests that there may be a 100 bp repetitive region within the NTSs of these two species in addition to the 250 bp periodicity.

Quantitative microdensitometry of the *D. mauritiana* and *D. simulans* *Eco*RI patterns (Fig. 4(a)) reveals a very small degree of overlap between the two species ($\sim 2\%$), indicating that the acquisition or loss of *Eco*RI sites has occurred in the majority of NTS copies of one or other of the species, since species divergence.

(c) *Concerted evolution of internal transcribed spacer copies*

The transcribed region of the rDNA of each species was investigated by reprobing nitrocellulose filters (shown in Fig. 2) with a cloned rDNA repeat known to lack an intron (INS$^-$, pDm238; Fig. 1). New bands of hybridization (i.e. the differences in pattern between Figs 2 and 5) correspond to the rDNA transcribed regions. A single major new band, lying between 2·25 and 2·45 kb is revealed in *Hae*III-digested DNA of all the species (Fig. 5(a)). The corresponding region, approximately 450 bp longer, is revealed in *Alu*I-digested rDNA of all species except *D. teissieri* and *D. orena*. This suggests that in five of the species, *Hae*III and *Alu*I sites flank a region that varies in length between the species. The 2·25 kb *Hae*III band and the 2·7 kb *Alu*I band of *D. melanogaster* include the internal transcribed spacer. It is probable, therefore, that the ITS varies in length between the species.

D. teissieri has gained an *Alu*I site within the majority of its ITSs, giving rise to fragments of 1·8 kb and 1·0 kb (Figs 3(b) and 5(b)). The *Hae*III and *Alu*I fragments of the ITS of *D. orena* differ by 600 bp rather than the common difference of 450 bp of the other species. This may be due to a loss of an *Alu*I site or a gain of an *Hae*III site that has occurred throughout the vast majority of the *D. orena* ITS units.

Unlike the considerable length heterogeneity of the NTSs both within and between species, each species has only one predominant ITS length. This length

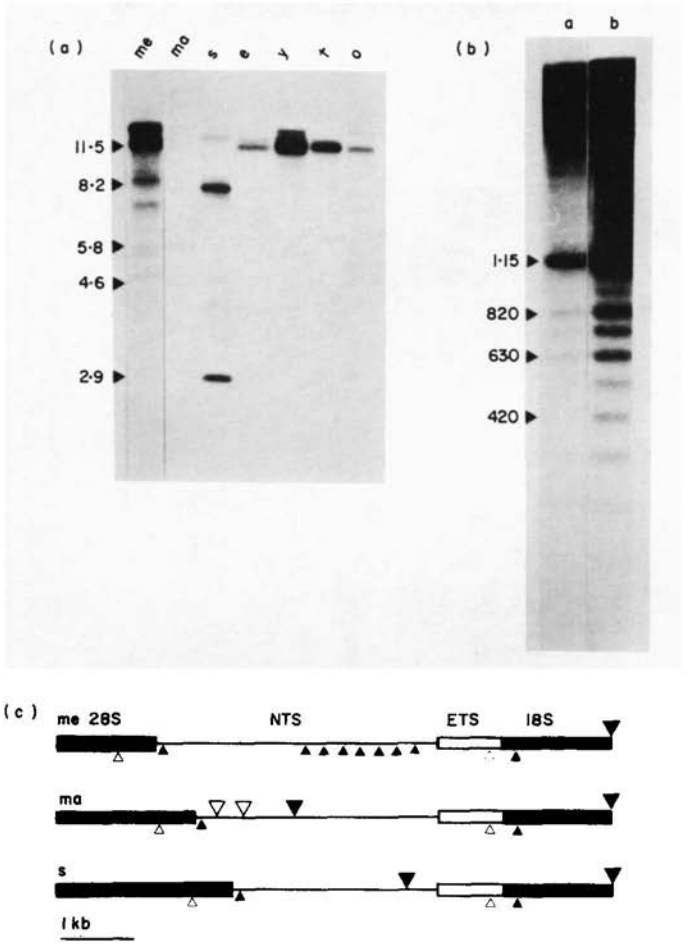


FIG. 4. (a) Autoradiogram of total female DNA from the 7 sibling species after digestion with *Eco*RI, fractionation on 0·8% agarose gel and hybridization to a ^{32}P-labelled clone of a rDNA repeating unit, pDm238. Each track contains the DNA from 1/3 of the same 8-fly extraction as used in Fig. 2. Sizes are in kb. The track containing *D. melanogaster* DNA was exposed for a third of the time used for the other tracks. (b) Autoradiogram of total male/female mixed DNA from *D. mauritiana* after digestion with *Eco*RI, fractionation on 2% agarose gel and hybridization with a ^{32}P-labelled NTS clone, pDm103HH2. Track b was exposed for 6 times as long as track a. Sizes are in base-pairs except for the topmost number, which is in kb. (c) Restriction maps of the NTSs of *D. melanogaster*, *D. mauritiana* and *D. simulans*. The restriction sites are: △, *Hae*III; ▲, *Alu*I; ▼, *Eco*RI; ▽, *Eco*RI sites probably not present in all the repeats.

varies between the species apart from *D. mauritiana* and *D. simulans*, which are closely related phylogenetically. Constancy in ITS length extends to the X and Y rDNA clusters in each species (see below).

The transcribed region, excluding the ITS, gives rise to many small *Alu*I fragments, which are conserved between all the sibling species (compare Fig. 3(a) and (b)). Such conservation is expected from the extensive data on 18 S ribosomal RNA and 28 S rRNA sequence conservation (see Gourse & Gerbi, 1980; Salim & Maden, 1981; Rae *et al.*, 1980).

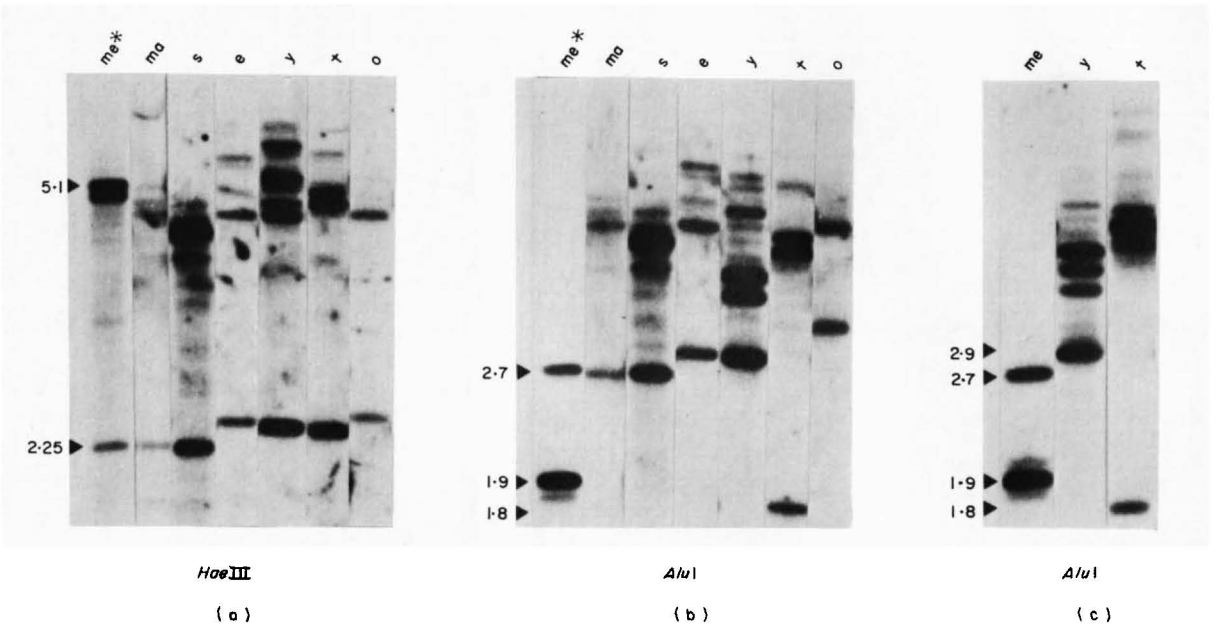


FIG. 5. (a) and (b) Autoradiogram of total female DNA from the 7 species after digestion with (a) *Hae*III or (b) *Alu*I, fractionation on 0·8% agarose gel and hybridization with a ^{32}P-labelled clone of a total rDNA repeat, pDm328. The same nitrocellulose filters as shown in Fig. 2 were used, after removal of the NTS probe. Sizes are in kb. An asterisk indicates a 5-fold shorter time of exposure for the autoradiogram. (c) Autoradiogram of total male DNA from *D. melanogaster*, *D. yakuba* and *D. teissieri* after digestion with *Alu*I, fractionation on a 0·8% agarose gel and hybridization with a ^{32}P-labelled clone of a total rDNA repeat, pDm238. Each track contains the DNA from half of an 8-fly extraction. Sizes are in kb.

(d) *Concerted evolution between X and Y rDNA clusters*

Y chromosome rDNA may be analysed by comparing hybridization patterns of XY male DNA and XX female DNA. *Alu*I digests of XY DNA, probed with the INS$^-$ total repeat, show that the internal transcribed spacer lengths for the seven species are the same for XY and XX DNA (*D. melanogaster*, *D. teissieri* and *D. yakuba* are shown in Fig. 5(c)). Some differences in the pattern of NTS fragments are observed between XX and XY DNA (e.g. in *D. teissieri* and *D. yakuba*, Fig. 5(b) and (c)). Such differences are unlikely to be due to polymorphism in the X chromosome, since both the male and female DNAs each derive from eight flies taken from an inbred laboratory stock. It can be seen that in *D. melanogaster* all of the NTSs of the Y rDNA, as in the X rDNA, contain *Alu*I sites. Also, the NTS probe has revealed no 250 bp fragment in *Alu*I-restricted male/female DNA of all the other species (Fig. 3(a)).

(e) *Species distribution of type I and type II introns*

Restriction of total female DNA of each species with *Hae*III and probing with a type I intron clone (pC225, Fig. 1) reveals homologous sequences in all seven of the species (6 are shown in Fig. 6). Some fragment lengths are shared between the species. For example, *D. melanogaster*, *D. simulans* and *D. mauritiana* have common lengths at 1·15 kb and 800 bp (Fig. 6). Most bands, however, are restricted to one or two of the species and some species have very few, if any, common fragment lengths (e.g. *D. orena* and *D. melanogaster*). *Eco*RI-digested DNA probed with the type I intron clone reveals several bands at positions that coincide with those observed with the INS$^-$ probe, suggesting that in these species some of the type I intron-like sequences are inserted in the rDNA units. The exact proportion is not known.

Type II intron–homologous sequences in the species were investigated using the cloned type II intron sequence peDm207C/D (Roiha & Glover, 1980). Hybridizing sequences are detected in *Hae*III-digested female DNA of all the species (data not shown). Probing *Eco*RI-digested male DNA of the species with the type II intron clone shows some bands at positions that coincide with bands revealed by the INS$^-$ rDNA unit probe, suggesting the presence of type II intron-like sequences within rDNA units of these species.

(f) *Concerted evolution of histone genes*

The five histone genes of *D. melanogaster* are clustered in about 100 tandem repeating units (Lifton *et al.*, 1977). In wild-type Oregon R there are two major repeat lengths of 5·0 kb and 4·8 kb due to differences in length of one of the spacers. The overall organization is complicated by a series of large spacer sequences (Karp, 1979; Saigo *et al.*, 1980; Kedes, 1979; Hentschel & Birnsteil, 1981).

The histone gene sequences of the species were analysed with the cDm500 clone of *D. melanogaster* (Lifton *et al.*, 1977). This clone contains almost two 4·8 kb histone repeats, each containing the five histone genes. *Bam*HI-restricted total DNA of the species was probed with a 1·7 kb *Hae*III fragment of cDm500 (Fig. 7(a)). *Bam*HI cuts the histone repeat of *D. melanogaster* once to give bands at

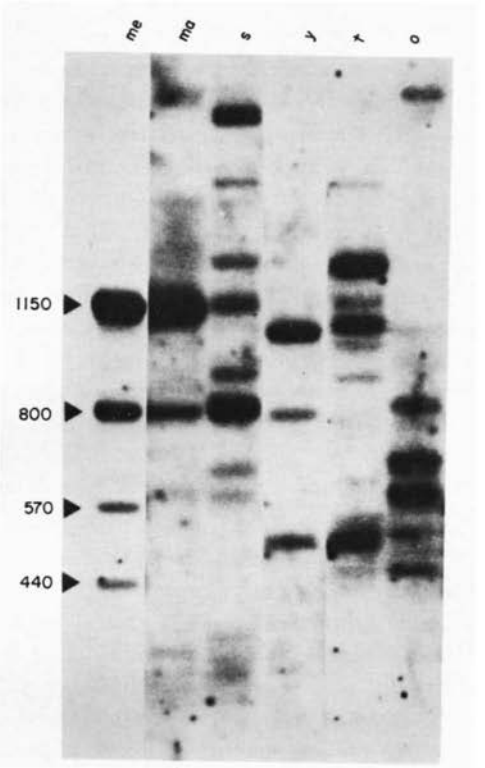


FIG. 6. Autoradiogram of total female DNA from 6 of the sibling species after digestion with *Hae*III and hybridization with a ^{32}P-labelled intron clone, pC225. Each track contains the DNA from an 8-fly extraction. The track containing *D. melanogaster* DNA was exposed for a 5-fold shorter time than the other tracks. Sizes are in bp.

4·8 kb and 5·0 kb corresponding to two major repeat lengths (Lifton *et al.*, 1977). Similar fragment sizes of 4·8 to 5·2 kb representing the total repeat length are observed in all species except *D. mauritiana*, which has a fragment smaller by 300 bp (Fig. 7(a)). This is due to the presence of an extra *Bam*HI site, since *Hin*dIII digests show *D. mauritiana* to have the same total histone repeat length as *D. simulans* (Fig. 7(b)). This extra *Bam*HI site is present in the majority, if not all, of the repeats of *D. mauritiana*.

Mapping the positions of the *Ava*I sites relative to the *Hin*dIII sites (Fig. 7(b)) shows conservation of these sites in the five species examined (see the legend to Fig. 7 for a full interpretation of the autoradiogram). In contrast to the results obtained using *Ava*I and *Bam*HI, *Hae*III digests reveal extensive divergence between the species (Fig. 8(a)), excepting *D. mauritiana* and *D. simulans*, which have identical patterns.

The positions of the *Hae*III sites were mapped relative to the *Bam*HI site and aligned with the *Hae*III restriction map of cDm500 (Goldberg, 1979; Fig. 8(b)). The overall shift of bands 300 bp to the left in *D. mauritiana* and *D. simulans* (relative to *D. melanogaster*) is due to the presence of an extra *Bam*HI site about 300 bp to

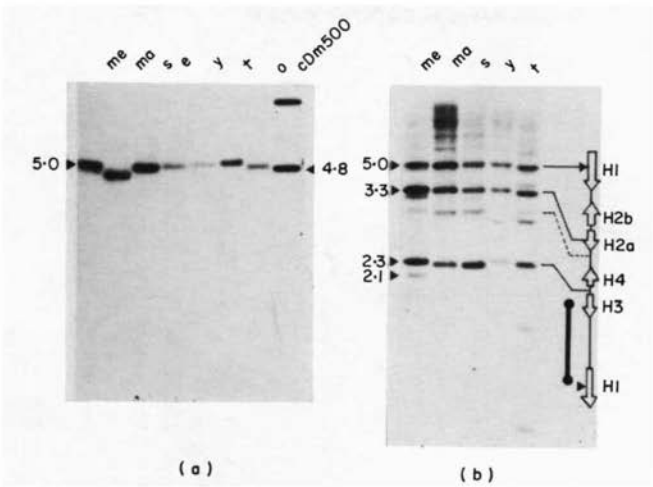


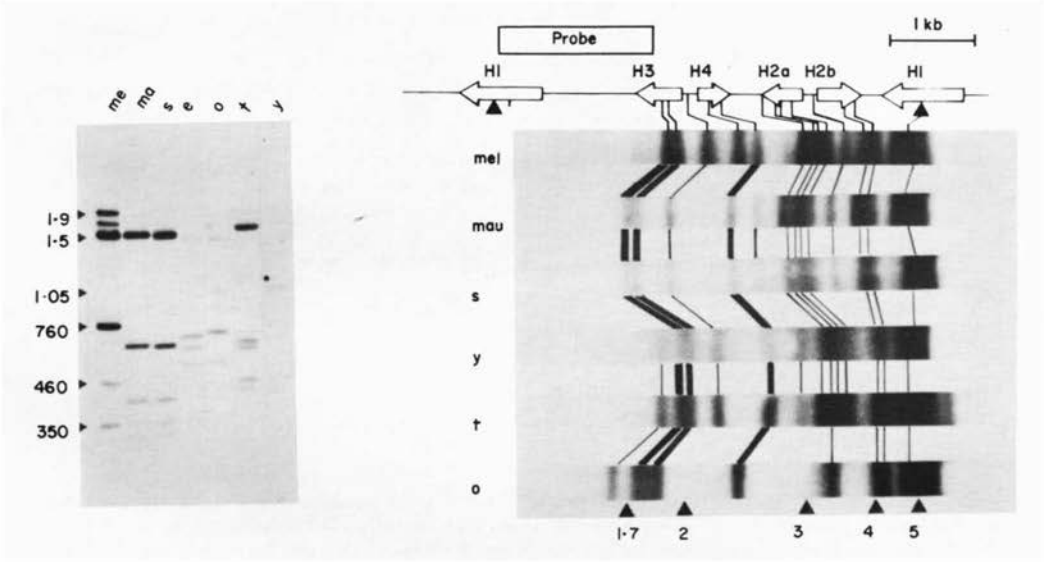
FIG. 7. (a) Autoradiogram of total DNA from the 7 species after digestion with *Bam*HI, fractionation on 1% agarose gel and hybridization to a ^{32}P-labelled histone *Hae*III fragment probe (see below). *Bam*HI-digested cDm500, hybridized to the same probe, is shown on the right for comparison. Sizes are in kb. (b) Autoradiogram of total DNA from 5 of the sibling species after digestion with *Hin*dIII, partial digestion with *Ava*I and hybridization to the ^{32}P-labelled 1·7 kb *Hae*III fragment of cDm500. Only *Ava*I/*Hin*dIII fragments that contain the large non-transcribed spacer at one end are revealed by the probe, so that the distance of the *Ava*I sites from the *Hin*dIII site can be mapped directly. A map of a *D. melanogaster* histone repeat is shown on the right. Transcribed regions are shown as open arrows, which point in the direction of transcription (Lifton *et al.*, 1977). The *Hae*III fragment probe is indicated by a vertical filled bar, and the position of the *Hin*dIII sites are shown by filled arrows. The site on the map to which each band on the autoradiogram corresponds is shown by a connecting line. The broken line indicates a site that is cleaved by a contaminant of the *Ava*I. Minor bands above 5 kb are due to the *Hin*dIII digestion being incomplete. Minor bands below 5 kb in the *D. teissieri* track are due to extra *Hin*dIII sites within the repeat that have been cleaved only partially. Minor bands in *D. melanogaster* 200 bp below the major bands derive from the 4·8 kb repeat.

the right of the position of the *Bam*HI site of *D. melanogaster*. *D. simulans* has only one *Bam*HI site in each repeat unit (Fig. 7(a)) and has therefore lost the site that is present in *D. melanogaster*. This may also account for the leftward shift of the *D. orena* fragments.

A significant feature to emerge from this comparison is that variation in *Hae*III restriction sites occurs in both the genes and spacers. For example, *D. mauritiana* and *D. simulans* lack a *Hae*III site that is present in the H4 coding region of *D. melanogaster* (Fig. 8(b)). The last three base-pairs of the *Hae*III sites (G-G-C-C) in the coding regions code for the amino acid alanine (Goldberg, 1979). Thus, the first G and the last C of the *Hae*III site are both wobble bases. This probably accounts for tolerance to changes in the *Hae*III sites of the coding region. It is also clear that most of the restriction site changes have spread to the majority, if not all, of the repeats in the histone array of each species (Fig. 8(a) and (b)).

4. Discussion

Many multigene families exhibit the phenomenon of concerted evolution. This is the observation that all the members of a family in a species share a feature (e.g. restriction site) that distinguishes them from their counterparts in another species.



(a) (b)

FIG. 8. (a) Autoradiogram of total DNA from the 7 species after digestion with *Hae*III and hybridization with a ^{32}P-labelled histone clone, cDm500. Sizes are in kb (top 3) or bp (bottom 3). The *D. melanogaster* strain used in this experiment was founded from a collection made in the Ivory Coast and shows a band at 1·5 kb in addition to the 1·7 and 1·9 kb bands expected from the 4·8 kb and 5·0 kb repeats, respectively. The 1·5 kb band is not seen in Canton-S histone DNA and may be due to a shorter spacer or an extra *Hae*III site. (b) Autoradiogram of total DNA from 6 of the species after complete restriction with *Bam*HI, partial digestion with *Hae*III and hybridization with a ^{32}P-labelled 1·7 kb histone *Hae*III fragment of cDm500. A map of a histone repeating unit of *D. melanogaster* is shown above the autoradiograms. The scale of the map is shown at the top right. Transcribed regions are indicated by open arrows, which point in the direction of transcription. The origin of the *Hae*III fragment is shown above the map. Numbers below the autoradiogram indicate sizes of the gel fragments in kb. Filled arrows on the map indicate the positions of the *Bam*HI sites. *Hae*III sites are shown as vertical lines and are connected to the corresponding bands in the *D. melanogaster* autoradiogram. Lines drawn between the bands of the different species indicate fragments at similar relative positions. The lines are shown thicker for the sites that are clearly conserved between all the species.

Many of the high molecular weight bands are not clearly resolved, so that lines connecting these bands between species are tentative.

The data relating to this phenomenon in families of globin (Jeffreys, 1982), histones (Kedes, 1979; Hentschel & Birnsteil, 1981), immunoglobulins (Baltimore, 1981), rDNA (Fedoroff, 1979; Dover & Coen, 1981) and many other families (Dover *et al.*, 1982; Long & Dawid, 1980) are reviewed, as indicated. Concerted evolution is an unfortunate term, in that it suggests a process of parallel evolution for all members of a family. Rather, it is an observation of family homogeneity that is an outcome of processes in the genome that are able to progressively fix a variant member in the family. Both unequal exchange of sequences (Smith, 1976) and gene conversion (Scherer & Davis, 1980) have been proposed as homogenizing mechanisms (see below). It is clear that homogenization occurs in many families that differ in copy-number, genomic dispersion or function (Dover *et al.*, 1982).

The data presented above on rDNA differences between the seven sibling species of the *D. melanogaster* subgroup reveal concerted evolution in several aspects. For example, in the *D. melanogaster* lineage alone, all the copies of the 250 bp region of internal repetition in each of the NTSs have been replaced by a variant member sensitive to *Alu*I (Fig. 4(c)). Similarly, variants diagnostic for each of the species in the subgroup, which are sensitive to *Eco*RI and *Alu*I, have become fixed in the copies of the NTS and ITS, respectively. Furthermore, each species is characterized by particular lengths of NTS and ITS. Fixation within species has also been found for variant sites in the coding and non-coding regions of the histone gene family. We have assessed what might be the evolutionary forces responsible for these patterns of variation, in the following manner.

(a) *Relative rates of sequence change and sequence fixation*

In order to illustrate the rate of fixation (concerted evolution) in relation to the rate of sequence change within different regions of the rDNA and histone families, we have constructed a phylogeny that is based on the extents to which each region of a family can be used to discriminate between the species (Fig. 9). For example, regions such as the coding sequences of the rDNA, which are homogeneous for the same sequences in all seven species, are placed at level F. At the other extreme, level A represents the within-species polymorphism for rDNA NTS length and copy-number of the different lengths (Coen *et al.*, 1982). An intermediate level, for example C, is illustrated by the length of the rDNA ITS, the variation of which cannot discriminate between *D. mauritiana* and *D. simulans* (Fig. 5). The histone *Hae*III sites span several levels, since there are insufficient data to assign individual levels to the various regions within the histone repeat unit.

In this phylogeny, we are able to place the minimum rates at which fixation has proceeded for the rDNA and histone units at level B. In order to explain how we arrived at these minimal rates, we make the following points.

The rate of evolution of each region of a family can be partitioned into its separate components. The rate of evolution in a multiple-copy family can be considered to be an outcome of the rate of change and the rate of fixation.

The rate of change is understood as the rate of production of effectively neutral mutations that are able to enter the homogenization process and become fixed. It is clear that the rates of change within different regions of the rDNA unit are different. This might be due to differences in the rate of mutation, or to selective constraints on the number and types of mutation that are permitted to proceed and enter the homogenization process. Given the biological function of the coding region, it is reasonable to suggest that the latter is the case for this region (level F). The placing of ITS length at level C suggests that either the rate of change of ITS length is slower than that of the NTS (level A), possibly due to the absence of an internal region of repetition, or that there are selective constraints on changes in the length of the ITS transcribed spacer.

The rate of fixation of a neutral variant unit in a population would depend on:

(1) the stochastic rate of homogenization of a variant unit in an array on a chromosome;

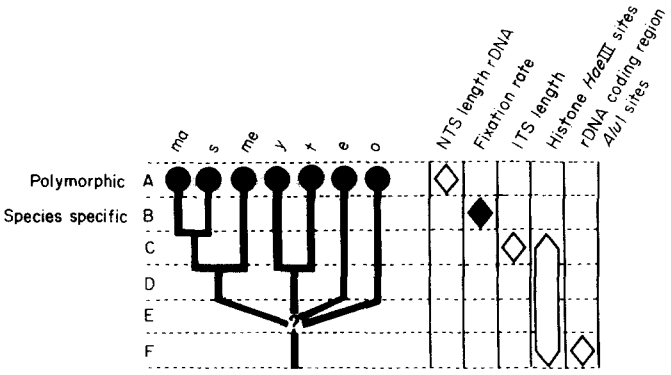


FIG. 9. Phylogeny of the 7 sibling species showing the relative rate of change in various regions of DNA and the rate of fixation of the rDNA and histone arrays (for definitions of rates of change and fixation, see the text). Rates of change are shown as open diamonds and the rate of fixation as a filled diamond. For rates of change, each feature of the DNA is placed at a level in the phylogeny according to the species classification it provides. For example, the rDNA internal transcribed spacer length has been placed at level C, since it is indistinguishable between *D. mauritiana* and *D. simulans* but is different for the other species. Similarly, NTS length is placed at level A, since the patterns of NTS length and abundance are highly polymorphic. The rate of fixation is placed at level B, since fixation has occurred between the 2 most closely related species, *D. mauritiana* and *D. simulans* (see the text for a full explanation). This phylogeny is consistent with those based on polytene chromosome inversions (Lemeunier & Ashburner, 1976) and abundant DNA families (Strachan *et al.*, 1982).

(2) the stochastic rate of fixation of the chromosomal array in a population;

(3) the rate of "transfer" of the variant units between chromosomes.

The dynamics of (1) have been examined theoretically with respect to the homogenizing potentials of unequal exchange and gene conversion by Smith (1976), Ohta (1980,1982), Birky & Skavaril (1976) and Nagylaki & Petes (1982). In addition, the dynamics of (2) are described by Kimura & Ohta (1979) and Ohta (1980). A detailed consideration of all three processes is given by Dover *et al.* (1982) (and see below).

For a given rate of fixation, there will be a certain time required for a newly arisen variant repeat to become fixed and replace the original repeats. Sufficient time has elapsed since the divergence of the two most closely related species, *D. mauritiana* and *D. simulans*, for variant repeats to become fixed in both the histone and rDNA arrays (as revealed by *Bam*HI and *Eco*RI sites). The minimum estimate for the rate of fixation in both families therefore corresponds to level B. Features of rDNA and histone DNA that are below level B (levels C to E) are also homogeneous in the arrays. This indicates that most regions of a repeat are being co-homogenized at a relatively fast rate (level B). The mechanism by which this might be achieved is described below.

Although it is probable that, in general, the rate of fixation lies within the range of level B, it is clear that the particular rate for any variant may be affected by selection, population size and any interaction of the variant with the homogenizing mechanism itself (i.e. unequal exchange or gene conversion). Before we can fully discuss the relative contributions of the three processes ((1), (2) and (3) defined above), to the rate of fixation we need to assess the extent to which homogenization

occurs between arrays on separate chromosomes; and also the fate of intervening sequences during the fixation process.

(b) *Homogenization between arrays*

Comparison of X and Y chromosome rDNA shows that homogenization has occurred in parallel between the two clusters. For example, in *D. melanogaster* both the X and Y chromosome NTSs contain the 250 bp-spaced *Alu*I sites. Similarly, the ITS lengths are the same for X and Y chromosome rDNA for each of the seven species. For arrays that are distributed between sex chromosomes, it is clear that fixation of a variant in both arrays cannot be due to the processes (1) and (2) alone but that a transfer of information between chromosomes (3) is required. We have recently been able to demonstrate, at the molecular and cytological levels, an exchange of sequences between X and Y rDNA arrays in *D. melanogaster* (our unpublished results). Homogenization between primate rDNA arrays also requires interchromosome transfer (Arnheim *et al.*, 1980).

There are, however, features that distinguish the X and Y rDNA of *D. melanogaster*. For example, the Y chromosome rDNA from Oregon R stocks does not contain any type I insertions (Tartof & Dawid, 1976; Wellauer *et al.*, 1978), and distinctive features of the Y chromosome rDNA have been found in the 18 S coding region (Yagura *et al.*, 1979) and in the NTS (Coen *et al.*, 1982; Endow & Glover, 1979). This implies that the rate of homogenization between X and Y arrays is less than the rate within each array. Alternatively, there may be selective adaptation of X and Y rDNA arrays that maintains their differences.

(c) *The behaviour of introns during homogenization*

Sequences homologous to both type I and type II introns are present in the seven sibling species. Insertion sequences homologous to the type I intron of *D. melanogaster* have been found also in the more distantly related *Drosophila* species *D. virilis* and *D. hydei* (Barnett & Rae, 1979; Renkawitz-Pohl *et al.*, 1980). Bearing in mind the age of the intron sequences and the observation of extensive sequence replacement in the rDNA array, one would expect a repeat containing or lacking an intron to have run to fixation throughout the rDNA. The presence in *D. melanogaster* of introns in a proportion of the rDNA units, taken in conjunction with the evidence for intron mobility (Rae *et al.*, 1980; Rae, 1981; Roiha *et al.*, 1981), suggests that intron sequences might be continually re-inserting into newly homogenized rDNA arrays (Dover & Coen, 1981).

(d) *The mechanism of homogenization*

Unequal exchange and gene conversion have been proposed as homogenizing mechanisms for tandem and interspersed families (for references, see Smith, 1976; Ohta, 1980; Baltimore, 1981; Nagylaki & Petes, 1982; Brown & Dover, 1981; Dover *et al.*, 1982; Fedoroff, 1979; Scherer & Davis, 1980). The unequal exchange model predicts fluctuations in copy-number. Therefore, unequal exchange within the 250 bp repetition of the NTS would result in variation in NTS length based on multiples of 250 bp. In addition, if unequal exchange occurs at a stagger of the total

rDNA repeats, then variation in copy-number would occur for many of the individual lengths produced by unequal exchange within the 250 bp repetition.

In Figure 10, we illustrate the two periodicities of unequal exchange and their predicted outcomes. Such variations in NTS length and copy-number have been observed within and between different chromosomal arrays in populations of *D. melanogaster*, *D. yakuba*, *D. teisseiri* and *D. hydei* and other Diptera (Coen *et al.*, 1982; Coen & Dover, unpublished results; Endow & Glover, 1979; Kunz *et al.*, 1981; Schäfer *et al.*, 1981). This suggests that unequal exchange is occurring at the two levels (between 250 bp repeats and between total rDNA units) and is rapid on an evolutionary time-scale. We show here that variation in NTS length both within and between species is also based on a 250 bp interval (Fig. 2(c)).

A prediction of relatively frequent unequal exchange at two different levels would be the homogenization of variants at these same two periodicities, and also lead to the co-homogenization of all regions within the total rDNA repeat unit (see above and Fig. 9). Homogenization at the two levels is observed for the 250 bp variants in the NTS of *D. melanogaster*. The NTS of *D. melanogaster* is known to form an uninterrupted heteroduplex with the NTS of *D. yakuba* (Tartof, 1979). In spite of this extensive matching of NTS sequences between the species, *Alu*I sites are located at regular 250 bp intervals throughout the NTSs of *D. melanogaster* but not in those of *D. yakuba*. Thus in the *D. melanogaster* lineage, a process of replacement by an *Alu*I-sensitive variant has occurred at the two different levels of organization. First, a variant 250 bp repeat containing an *Alu*I site has spread through the region of repetition within the NTS, and secondly, an NTS containing these variant "*Alu*I" repeats has become fixed throughout the rDNA. Such schemes would also explain the NTS sequence variation observed in other species of *Drosophila* (Rae *et al.*, 1981).

(e) *The rate of fixation*

In order to determine whether unequal exchange (1) and drift (2) (see above) are sufficiently rapid to ensure the fixation of different rDNA variants in the two species *D. melanogaster* and *D. simulans*, we have made the following calculations based on the considerations of Tartof (1974), Smith (1976) and Ohta (1980).

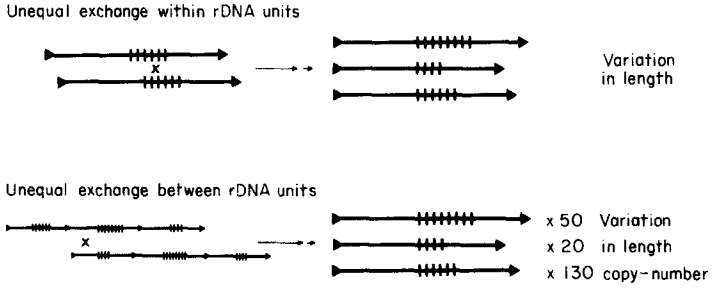


FIG. 10. Top: Unequal exchange at the periodicity of an internal region of repetition within a larger repeating unit (cf. rDNA: Fig. 1). This generates differences in length of the longer repeat by multiples of the internal repeat length. Bottom: Unequal exchange at the periodicity of the longer repeat unit (e.g. total rDNA unit) generates variation in copy-number of the different lengths of the unit.

Assuming that only unequal sister chromatid exchange occurs, i.e. within a single chromosomal array, and that the number of repeats is constrained to lie within 10% of their mean, then the time required for the spread of a neutral variant unit to 200 copies of an array will be of the order of 10^3 to 10^4 crossover generations. Given a rate of unequal exchange of approximately 3×10^{-4} per generation (see above; and Coen *et al.*, 1982; Frankham *et al.*, 1980), the time required for homogenization within an array is of the order of 10^7 generations. The average time taken for a chromosome, carrying an array fixed for a neutral variant, to spread throughout the population is $4N_e$ generations, where N_e is the mean effective population size. Thus, if N_e is 10^7 or less, the time taken for fixation would be limited by the rate of homogenization and hence be of the order of 10^7 generations or about 10^6 years, assuming ten generations per year.

Estimates of the time of divergence of *D. melanogaster* and *D. simulans* vary within an order of magnitude. If the most recent estimate of 10^6 generations is to be believed (Leigh Brown & Ish-Horowicz, 1981), then the time might be insufficient for fixation by unequal exchange and drift alone. The lack of time becomes more critical in considering the fixation of species-diagnostic variants of the rDNA and histone families in the two most closely related species, *D. simulans* and *D. mauritiana* (level B in Fig. 9). As we explain above, the minimal estimate for the rate of fixation is given by the time of separation of these two most recently diverged species within the subgroup. In order to determine whether the rate is indeed faster than level B, we are currently examining the rDNA, histone and non-coding DNA variation in a newly described species, *Drosophila sechellia* (Tsacas & Bachli, 1981) that is phylogenetically very close to *D. simulans* and *D. mauritiana*. It may indeed have arisen since divergence of the latter two species and hence divide the interval between them (see Fig. 9).

If the processes of unequal exchange and drift are insufficient to account for the fixation of variants in the rDNA and histone families between *D. simulans* and *D. mauritiana*, in what way can fixation be accelerated? There are two ways in which this might be achieved. One is by selection and the other is by a process that we have termed molecular drive (Dover *et al.*, 1982). The effect on phenotype of one variant member within a family of sequences is probably minimal. Hence selection might influence the evolutionary progression of a variant only after it has increased in frequency by drift or drive. Molecular drive is the consequential increase in frequency of any sequence that is able to transfer between chromosomes in a biased manner due to a continual process of non-Mendelian transmission. We discuss this issue in more detail in the accompanying paper, in which we describe the rates of fixation within families of large copy-number and wide karyotype dispersion (Strachan *et al.*, 1982). The involvement of sequence transposition and biased gene conversion in the inter-chromosome transfer process, and the extent to which they occur and are able to accelerate the process of fixation in multiple-copy families (molecular drive), is discussed in full by Dover *et al.* (1982).

We express our thanks to Francoise Lemeunier and her colleagues for the supply of the *Drosophila* species. The comments of Steven Brown, John Dallas, Bob Karp and Martin Trick are appreciated. This research has been supported by Science Research Council grant

numbers GR/A/6537 and GR/A/3758.1. One of us (E.S.C.) acknowledges receipt of a Science Research Council studentship.

REFERENCES

Arnheim, N., Krystal, M., Schnickel, R., Wilson, G., Ryder, O. & Zimmer, E. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 7323–7327.
Baltimore, D. (1981). *Cell*, **24**, 592–594.
Barnes, S. R., Webb, D. A. & Dover, G. A. (1978). *Chromosoma*, **67**, 341–363.
Barnett, T. & Rae, P. M. M. (1979). *Cell*, **16**, 763–775.
Birky, C. W. Jr & Skavaril, R. V. (1976). *Genet. Res. (Camb)*, **27**, 249–265.
Brown, S. D. M. & Dover, G. A. (1981). *J. Mol. Biol.* **150**, 441–466.
Brown, D. D., Wesink, P. C. & Jordan, E. (1972). *J. Mol. Biol.* **63**, 57–73.
Coen, E. S., Thoday, J. M. & Dover, G. A. (1982). *Nature (London)*, **295**, 564–568.
Dover, G. A. & Coen, E. S. (1981). *Nature (London)*, 731–732.
Dover, G. A., Brown, S. D. M., Coen, E. S., Dallas, J., Strachan, T. & Trick, M. (1982). In *Genome Evolution* (Dover, G. A. & Flavell, R. B., eds), pp. 343–374. Academic Press, London.
Endow, S. A. & Glover, D. M. (1979). *Cell*, **17**, 597–605.
Fedoroff, N. V. (1979). *Cell*, **16**, 697–710.
Flavell, R. B. (1982). In *Genome Evolution* (Dover, G. A. & Flavell, R. B., eds), pp. 301–324. Academic Press, London.
Ford, P. J. & Brown, D. D. (1976). *Cell*, **8**, 485–493.
Frankham, R., Briscoe, D. A. & Nurthen, R. K. (1980). *Genetics*, **95**, 727–742.
Glover, D. J. & Hogness, D. S. (1977). *Cell*, **10**, 167–176.
Goldberg, M. L. (1979). Ph.D. thesis, Stanford University.
Gourse, R. L. & Gerbi, S. A. (1980). *J. Mol. Biol.* **140**, 321–339.
Hentschel, C. C. & Birnsteil, M. L. (1981). *Cell*, **25**, 301–313.
Jeffreys, A. (1982). In *Genome Evolution* (Dover, G. A. & Flavell, R. B., eds), pp. 157–176. Academic Press, London.
Karp, R. W. (1979). Ph.D. thesis, Stanford University.
Kedes, L. H. (1979). *Annu. Rev. Biochem.* **48**, 837–870.
Kimura, M. & Ohta, T. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 4001–4005.
Kunz, W., Peterson, G., Renkawitz-Pohl, R., Glätzer, K. H. & Schäfer, M. (1981). *Chromosoma*, **83**, 145–158.
Lauer, J., Shen, C. J. & Maniatis, T. (1980). *Cell*, **20**, 119–130.
Leigh Brown, A. J. & Ish-Horowicz, D. (1981). *Nature (London)*, **290**, 677–682.
Lemeunier, F. & Ashburner, M. (1976). *Proc. Roy. Soc. ser. B*, **193**, 275–294.
Lifton, R. P., Goldberg, M. L., Karp, R. W. & Hogness, D. S. (1977). *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1047–1051.
Long, E. O. & Dawid, I. B. (1979). *Nucl. Acids Res.* **7**, 205–215.
Long, E. O. & Dawid, I. B. (1980). *Annu. Rev. Biochem.* **49**, 727–764.
Moore, G. P., Scheller, R. H., Davidson, E. H. & Britten, R. J. (1978). *Cell*, **15**, 649–660.
Nagylaki, T. & Petes, T. D. (1982). *Genetics*, in the press.
Ohta, T. (1980). *Evolution and Variation of Multigene Families, Lecture Notes in Biomathematics*, vol. 37. Springer-Verlag, Berlin.
Ohta, T. (1982). *Genetics*, in the press.
Pan, J., Elder, J. T., Duncan, C. & Weissman, S. M. (1981). *Nucl. Acids Res.* **9**, 1151–1170.
Rae, P. M. M. (1981). *Nucl. Acids Res.* **9**, 4997–5010.
Rae, P. M. M., Kohorn, B. D. & Wade, R. P. (1980). *Nucl. Acids Res.* **8**, 3491–3504.
Rae, P. M. M., Barnett, T. & Murtif, V. L. (1981). *Chromosoma*, **82**, 637–655.
Renkawitz-Pohl, R., Glatzer, K. H. & Kunz, W. (1980). *Nucl. Acids Res.* **8**, 4593–4611.
Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–251.
Roiha, H. & Glover, D. M. (1980). *J. Mol. Biol.* **140**, 341–355.

Roiha, H., Miller, J. R., Woods, L. C. & Glover, D. M. (1981). *Nature (London)*, **290**, 749–753.
Saigo, K., Millstein, L. & Thomas, C. A. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 815–827.
Salim, M. & Maden, B. E. H. (1981). *Nature (London)*, **291**, 205–208.
Schäfer, M., Wyman, A. R. & White, R. (1981). *J. Mol. Biol.* **146**, 179–199.
Scherer, S. & Davis, R. W. (1980). *Science*, **209**, 1380–1384.
Smith, G. P. (1976). *Science*, **191**, 528–535.
Strachan, T., Coen, E. S., Webb, D. A. & Dover, G. A. (1982). *J. Mol. Biol.* **158**, 37–54.
Tabak, H. F. & Flavell, R. A. (1978). *Nucl. Acids Res.* **5**, 2321–2332.
Tartof, K. (1974). *Cold Spring Harbor Symp. Quant. Biol.* **38**, 491–500.
Tartof, K. (1979). *Cell*, **17**, 607–614.
Tartof, K. D. & Dawid, I. B. (1976). *Nature (London)*, **263**, 27–30.
Tsacas, L. & Bachli, G. (1981). *Revue fr. Ent. (N.S.)*. **3**, 146–150.
Wahl, G. M., Stern, M. & Stark, G. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 3683–3687.
Wellauer, P. K., Dawid, I. B. & Tartof, K. D. (1978). *Cell*, **14**, 269–278.
Yagura, T., Yagura, M. & Muramatsu, M. (1979). *J. Mol. Biol.* **133**, 533–547.
Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. & Wilson, A. C. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 2158–2162.

Edited by S. Brenner