Evolution of Ribosomal RNA Gene Copy Number on the Sex Chromosomes of *Drosophila melanogaster*[1]

Eva M. S. Lyckegaard and Andrew G. Clark

Department of Biology and Genetics Program, Institute of Molecular Evolutionary Genetics, Pennsylvania State University

A diverse array of cellular and evolutionary forces—including unequal crossing-over, magnification, compensation, and natural selection—is at play modulating the number of copies of ribosomal RNA (rRNA) genes on the X and Y chromosomes of *Drosophila*. Accurate estimates of naturally occurring distributions of copy numbers on both the X and Y chromosomes are needed in order to explore the evolutionary end result of these forces. Estimates of relative copy numbers of the ribosomal DNA repeat, as well as of the type I and type II inserts, were obtained for a series of 96 X chromosomes and 144 Y chromosomes by using densitometric measurements of slot blots of genomic DNA from adult *D. melanogaster* bearing appropriate deficiencies that reveal chromosome-specific copy numbers. Estimates of copy number were put on an absolute scale with slot blots having serial dilutions both of the repeat and of genomic DNA from nonpolytene larval brain and imaginal discs. The distributions of rRNA copy number are decidedly skewed, with a long tail toward higher copy numbers. These distributions were fitted by a population genetic model that posits three different types of exchange events—sister-chromatid exchange, intrachromatid exchange, and interchromosomal crossing-over. In addition, the model incorporates natural selection, because experimental evidence shows that there is a minimum number of functional elements necessary for survival. Adequate fits of the model were found, indicating that either natural selection also eliminates chromosomes with high copy number or that the rate of intrachromatid exchange exceeds the rate of interchromosomal exchange.

Introduction

The steady-state distribution of the number of copies of a gene in a multigene family depends on a variety of forces operating at both the cellular and population levels. There is a rich history of research on the ribosomal DNA (rDNA) array of *Drosophila melanogaster,* and consideration of the evolution of the rDNA array requires understanding of these results. The 28S and 18S ribosomal RNA (rRNA) genes of *D. melanogaster* are distributed into two regions, known as nucleolus organizers (NO), on the sex chromosomes (Ritossa 1976). Estimated numbers of copies on each chromosome are 100–240 in wild-phenotype laboratory stocks (Long and Dawid 1980). Low copy number is associated with the bobbed (*bb*) phenotype, characterized by delayed development, abdominal etching, and thin, short bristles. At a copy number below ~15% of the wild-type number, embryonic lethality results (Long and Dawid

1. Key words: rDNA, tandem array, population genetic theory, *Drosophila melanogaster.*

Address for correspondence and reprints: Eva M. S. Lyckegaard, Department of Biology and Genetics Program, Institute of Molecular Evolutionary Genetics, 208 Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802.

Mol. Biol. Evol. 8(4):458–474. 1991.

0737-4038/91/0804-0005$02.00

1980). The rDNA unit has been cloned and sequenced (Glover and Hogness 1977; Tautz et al. 1988), revealing a structure with an intergenic sequence (IGS) of 3,632 bp, an external transcribed spacer (ETS) of 864 bp, the 18S unit (1,995 bp), and a 28S unit (3,945 bp). Between the 18S and 28S genes is an internal transcribed spacer (ITS) that encodes a 5.8S and a 2S rRNA. The 28S unit is frequently interrupted by either type I or type II insertion sequences. Type I sequences interrupt the 28S unit in ~60% of the X-chromosome copies and are 0.5–5.35 kb in length (Jakubczak et al. 1990). Type II inserts occur in ~15% of the 28S rDNA units on both the X and Y chromosomes and are 1.5–3.6 kb in length (Wellauer et al. 1978; de Cicco and Glover 1983). Transcripts of the interrupted genes can be detected, but they occur at very low levels and fail to produce mature rRNA, even in *bobbed* mutants (Kidd and Glover 1981; Long et al. 1981).

Several classes of exchange events are responsible for variation in copy number. Compensation refers to differential replication of rDNA such that the rDNA content of XX and XO flies is the same, indicating a twofold-higher level of amplification in the XO flies (Tartof 1971). Compensation is a purely somatic phenomenon and is relevant to the evolution of the rDNA array only to the extent that it alters the fitness of flies with low copy number. Magnification refers to the reversion of the *bb* phenotype mediated by increased germ-line copy numbers. Magnification is most frequently observed among the gametes of males that are low in rDNA on both sex chromosomes. X-Y translocations reveal that part of the long arm of the Y, a part distinct from NOR, is necessary for magnification in males and that females that have this part of the YL also magnify (Komma and Endow 1987). Magnification results in amelioration of the *bb* phenotype, so active genes are involved, but whether genes lacking the insertion sequences are preferentially amplified remains controversial (de Cicco and Glover 1983; Terracol and Prud'homme 1986). The dramatic changes in copy number that are associated with magnification appear to occur more frequently when there is a physiological demand for rRNA.

In the present study we estimate the number of copies of rDNA repeats, and the number of type I and type II inserts, in a series of X and Y chromosomes from a single sample of a natural population of *D. melanogaster*. The aim is to investigate, by analysis of the distributions of copy number, the evolutionary forces influencing the rDNA array. Copy number is altered by a variety of asymmetric exchange events, and the goal is to determine for these events the relative rates that are consistent with observed distributions of copy number. Unequal sister-chromatid exchange and interchromosomal unequal crossing-over result in products that have both higher and lower copy number. If these were the only forces at play, then the variance in copy number would grow without bound. The physiological importance of rRNA makes the array a target for natural selection acting at the organismal level. If there is an optimal copy number, and if fitness decreases as a quadratic function of the departure from this optimum, then copy number can be stabilized—and in fact has a normal distribution at equilibrium (Crow and Kimura 1970, pp. 294–296). Even in the absence of natural selection, intrachromatid exchange, which can occur when the array loops back onto itself, results in a reduction in copy number and can prevent the unbounded growth of the array (Walsh 1987).

Material and Methods

Origin of the *Drosophila* Lines

Drosophila melanogaster were collected at the Harner Farm peach orchard in Centre County, Pennsylvania, in August 1988. X-chromosome copy numbers were determined by crossing field-caught males to sc^4sc^8 females. Female offspring of this cross were heterozygous for the sc^4sc^8 rDNA deficiency, and all bear the paternal X chromosome. Y-chromosome variation was isolated by crossing wild males to virgin females bearing the $Df(1)bb^{1-158}$ *y* chromosome. The *yellow* male offspring, bearing the Xbb^{1-158} and the wild Y chromosome were collected for DNA extractions. Stocks bearing the $Df(1)bb^{1-158}$ *y* and sc^4sc^8 deficiencies were obtained from the Pasadena stock center (now at Indiana University). Total genomic DNA was isolated from adults, and RNA was removed by thorough RNase digestion according to the protocol of Clark and Lyckegaard (1988).

The equivalence of copy-number estimates over the two different deficiencies was verified by isolating and blotting the DNA from males bearing the sc^4sc^8 X deficiency and Y chromosomes that had also been tested over the Xbb^{1-158} deficiency. Finally, to determine whether the relative levels of polytenization of *Adh* and rDNA varied, we also isolated brain tissue and imaginal discs from ~200 larvae of each of two Y-lines and two X-lines. The genomic DNA of these nonpolytene tissues was extracted as described above.

DNA Slot-Blot Analysis

Replicated samples of DNA were loaded into a Bio-Dot SF slot-blot apparatus that focuses the DNA in thin lines on Zeta-Probe nylon membranes (Bio-Rad). The DNA samples were denatured in 0.4 M NaOH for 10 min and were neutralized by addition of an equal volume of 2 M NH_4OAc, pH 7. The denatured DNA was applied to randomized locations among the 48 slots/membrane. Each line was tested in replicates distributed on two different membranes. Four hundred microliters of 2 × SSC (1 × SSC = 0.15 M NaCl/0.015 M sodium citrate) was added to each well after the samples filtered through, and vacuum was applied until the sample wells were completely dry. The membrane was rinsed in 2 × SSC, was air-dried, and baked at 80°C for 1 h prior to hybridization. Membranes with a dilution series of the rDNA repeat, the type I and type II inserts, and brain and imaginal disc DNA were also prepared and quantified, in order to estimate absolute copy numbers.

Plasmid DNA

The membranes were hybridized with four different plasmids. The first plasmid, *p13E3* (Goldberg 1980) containing the *D. melanogaster Adh* gene (alcohol dehydrogenase; E.C. 1.1.1.1) in a 4.75-kb *Eco*RI fragment cloned into *pUC13*, served as a single-copy control for quantifying the total amounts of DNA bound to the membranes. The second plasmid, *pDmr.a51#1* (Dawid et al. 1978; Endow 1982), contains a complete 11.5-kb intron$^-$ ribosomal DNA repeat, from the X chromosome, cloned into *pACYC184*. The third and fourth plasmids, *pC24* (Long and Dawid 1979) and *p0.7kB* (Long et al. 1980), contained fragments of the type I and type II inserts cloned into *pBR322*. The plasmids were labeled with (α^{32}P)dCTP by nick-translation prior to hybridization.

Hybridization

The membranes were prehybridized at 65°C for 10 min with agitation in a prewarmed mixture of 1% bovine serum albumin, 1 mM ethylenediaminetetraacetate, 7% sodium dodecyl sulfate (SDS), 0.5 M sodium phosphate, pH 7.2. They were never allowed to dry completely after the first prehybridization. The prehybridization solution was removed and was replaced with the same solution and the denatured probe DNA. The hybridization continued at 65°C for 18 h with agitation. To remove nonspecifically bound probe after the hybridization, the membrane was washed at room temperature for 15 min each in 2 × SSC, 0.1% SDS; 0.5 × SSC, 0.1% SDS; and 0.1 × SSC, 0.1% SDS. The radiographic exposure was made with the moist membrane enclosed in a sealed plastic bag. A series of exposures was made for each hybridization, and the intensities of the bands on the resulting autoradiographs were quantified by computing the peak areas with scanning laser densitometry (LKB Ultroscan XL). Before each new hybridization, the previously used probe was removed by washing the membrane in 0.4 M NaOH at 65°C for 30 min and thereafter neutralizing with 0.1 × SSC, 0.5% SDS, 0.2 M Tris-HCl, pH 7.5, at 65°C for 30 min. A 24-h autoradiographic exposure was then done to verify the complete removal of the labeled probe. Subsequent probes were hybridized and assayed as described above.

Densitometry and Analysis

The resulting band intensities on the autoradiographs were quantified using scanning laser densitometry. Some autoradiographs had bands that spanned beyond the linear range of the film, so the analysis made use of all of the data by fitting the exposures to the full sensitometric curve of the film (Lyckegaard and Clark 1989). This was done by doing a logistic transformation [$D_{ijk} = \ln(p_{ijk}/(1 - p_{ijk}))$], where p_{ijk} is the band density scaled between 0 (unexposed) and 1 (saturated exposure) for replicate k of exposure j of line i and where D_{ijk} is the transformed band density. The following model was then fitted by least squares:

$$Q = \sum_i \sum_j \sum_k \{D_{ijk} - [\beta \log(t_j) + \alpha_{ik}]\}^2,$$

where β is a slope parameter for the sensitometric curve of the film (common to all lines and replicates), t_j is the exposure time, and α_{ik} is the intercept estimated separately for each replicate of each line. The estimates of β and α_{ik} that minimize Q were obtained numerically using a simplex algorithm (Press et al. 1986). Absolute copy numbers were calculated by linear regression using densities of the dilution series in the linear portion of the densitometric curve of the film. The ratio of the regression coefficient for genomic DNA and the cloned gene reflects the fraction of the genome that the cloned gene represents. This was converted to an absolute copy number by multiplying by the size of the *Drosophila* genome (170,000 kb) and dividing by the length of the cloned fragment (11.5 kb for rDNA, 0.8 kb for the type I insert, and 0.7 kb for the type II insert).

Theory and Simulations

To determine whether it was necessary to invoke natural selection as a force affecting the distribution of copy numbers and to estimate the relative rates of exchange

events, we fitted models of copy-number evolution to the data. The models, which are an extension of the work of Takahata (1981) and Walsh (1987), consider a diploid population with a multigene family that is undergoing sister-chromatid exchange, interchromosomal unequal crossing-over, and intrachromatid crossing-over (fig. 1).

Let x_i be the frequency of X chromosomes with i copies of the gene, let y_i be the frequency of Y chromosomes with i copies, let $z_{f,ij}$ be the frequency of female zygotes with i and j copies on the two X chromosomes, and let $z_{m,ij}$ be the frequency of male zygotes with i copies on the X chromosome and with j copies on the Y chromosome. Let sister-chromatid exchange occur at rate β_i per chromosome, allowing the rate to depend on i. Interchromosomal unequal crossing-over occurs at rate $\gamma_{i,j}$ between X chromosomes in females and at rate ω_{ij} between the X and Y chromosomes in males. Intrachromatid exchange occurs at rate δ_i. If a sister-chromatid exchange event occurs in a chromatid with i copies, it is assumed that the exchange occurs with a uniform distribution across the tandem array, so the products have a uniform distribution on $[1,2i-1]$. Similarly, a uniform distribution of alignments and locations of exchange events is assumed for interchromosomal unequal crossing-over and for intrachromatid crossing-over. In the latter case, the loop that is formed is excised, leaving a shorter array.

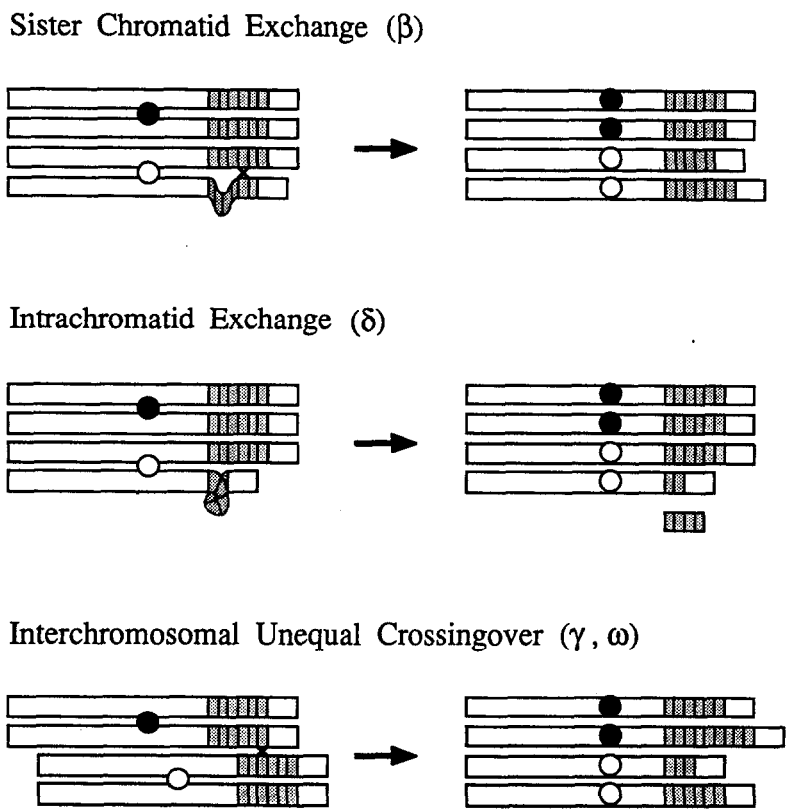

FIG. 1.—Three classes of exchange events, and parameters in model that describe their rates. Interchromosomal unequal crossing-over occurs between two X chromosomes at rate γ and occurs between an X and a Y chromosome at rate ω. For all exchange events, it is assumed that the chromosomes align in any possible register and that the probability of an exchange event is independent of copy number.

With random mating, the frequency of zygotes is $z_{f,ij} = x_i x_j$. After a zygote is formed, it is subjected to natural selection. Genotypes with i and j copies of the gene have fitness w_{i+j}, so

$$z'_{f,ij} = \frac{w_{i+j} z_{f,ij}}{\bar{w}}. \tag{1}$$

A similar equation gives the change in male genotype frequencies that is due to natural selection. The recursions that follow will also occur in pairs for each sex, with the exception of interchromosomal exchange. Letting $P_{j,k}$ be the probability that a chromosome with k copies ends up with j copies after a sister-chromatid exchange, we get

$$z''_{ij} = (1 - \beta_i)(1 - \beta_j) z'_{ij} + \sum_{\substack{k=1 \\ (j<2k)}}^{\infty} P_{j,k}(1 - \beta_i)\beta_k z'_{ik} + \sum_{\substack{k=1 \\ (j<2k)}}^{\infty} P_{i,k}(1 - \beta_j)\beta_k z'_{jk}. \tag{2}$$

Similarly, letting $R_{j,k}$ be the probability that a chromosome with k copies ends up with j copies after an intrachromatid exchange, we get

$$z'''_{ij} = (1 - \delta_i)(1 - \delta_j) z''_{ij} + \sum_{\substack{k=1 \\ (j<k)}}^{\infty} R_{j,k}(1 - \delta_i)\delta_k z''_{ik} + \sum_{\substack{k=1 \\ (j<k)}}^{\infty} R_{i,k}(1 - \delta_j)\delta_k z''_{jk}. \tag{3}$$

Finally, interchromosomal unequal crossing-over may take place. Let $Q_{j,k}$ be the probability that a genotype with a total of k copies on both chromosomes produces a gamete with j copies after an unequal exchange event. The probability of an interchromosomal unequal crossing-over is γ in females, so the recursion is

$$x'_i = \sum_{j=1}^{\infty} (1 - \gamma_{ij}) z'''_{ij} + \sum_{\substack{j=1 \\ (j+k)>i}}^{\infty} \sum_{k=1}^{\infty} Q_{i,j+k} \gamma_{jk} z'''_{jk}. \tag{4}$$

For males there is a similar pair of equations expressing the frequency of X and Y chromosome types after X-Y exchange events, which occur at a rate ω. Equations (1)–(4) form the recurrence system. For each set of parameters, an equilibrium distribution of copy numbers was determined by simulation. Exploratory runs were made to determine the influence that the parameters had on the steady-state distribution.

Iteration of equations (1)–(4) to equilibrium, for each step of a parameter estimation routine, would require an inordinate amount of computer time. With weak selection and low rates of exchange, three simplifications were made to speed the convergence at each step. First, the fitness function was assumed to be a ramp function, with individuals having $< w_1$ copies being given a fitness of 0, individuals with $> w_2$ copies being given a fitness of 1, and fitness being a linear function of copy number in individuals whose copy number lies between w_1 and w_2. Second, the rate of interchromosomal exchange is expected to be on the order of 10^{-4}/chromosome/generation (Williams et al. 1989), so the probability of both chromosomes in a diploid undergoing intrachromatid exchanges is very low. We can replace equations (2) and (3) with the appropriate equations that specify changes in chromosome frequency, and the genotype

frequencies can then be expressed as a product of chromosome frequencies after these two exchange events occur. Hence, equations (2) and (3) were replaced, respectively, with

$$x_i'' = (1 - \beta_i)x_i' + \sum_{\substack{j=1 \\ (i<2j)}}^{\infty} P_{i,j}\beta_j x_j' \tag{5}$$

and

$$x_i''' = (1 - \delta_i)x_i'' + \sum_{\substack{j=1 \\ (i<j)}}^{\infty} R_{i,j}\delta_j x_j''. \tag{6}$$

The third change that was made for the purposes of making the simulations practical was to split the distribution of copy numbers into bins of 10 copies. The rates that we estimate are then rates of transition from one bin to another. The near equivalence of the two recursion systems [i.e., eqq. (1)–(4) and eqq. (1) and (4)–(6)] was verified numerically for a few sets of parameters in the neighborhood of the final estimates.

With recurrence equations (1) and (4)–(6) the simplex method was used to minimize the χ^2 for the fit of this distribution to the observed distributions of copy number (Press et al. 1986). For each step of this process, the parameters were selected, and the copy number distribution was iterated to steady state. (It converges from a uniform distribution to its steady state at a rate that depends on the rates of exchange.) Starting each iteration of the stepping procedure with the observed distribution of copy numbers meant that convergence to the steady-state distribution was rapid. The χ^2 statistic was then calculated to test the goodness-of-fit between the observed distribution and the model. The simplex algorithm specified the rules for stepping the parameters to find those that yielded the best fit (i.e., smallest χ^2). Each distribution was fitted under various assumptions about which exchange events occur. When X-Y interchromosomal exchange was disallowed, the X- and Y-chromosomal arrays were fitted independently. Numerical work was performed in double precision on a SUN 3/50 workstation.

Results

Empirical Observations

For both the 96 X chromosomes and the 144 Y chromosomes, regression analysis of band densities obtained by densitometric scans of the slot blots produced estimates of the copy numbers of rDNA relative to the single-copy *Adh* gene. Two X and two Y chromosomes were also tested in a dilution series, along with a dilution series of the cloned insert of rDNA and type I and type II inserts. These blots were probed with the same four probes, and absolute amounts of DNA of each class were estimated by regression, using only points on the linear portion of the sensitometric curve of the film. These provided estimates of absolute copy numbers for two X and two Y chromosomes. The two estimates of absolute copy number allowed the set of relative copy numbers to be placed on a scale of absolute copy numbers. Figures 2–4 show these estimates for the X-chromosomal rDNA, the Y-chromosomal rDNA, and the type I and type II inserts. Because the type I inserts are present at sites other than the sex chromosomes, the accuracy of the estimates may be compromised by copy-number

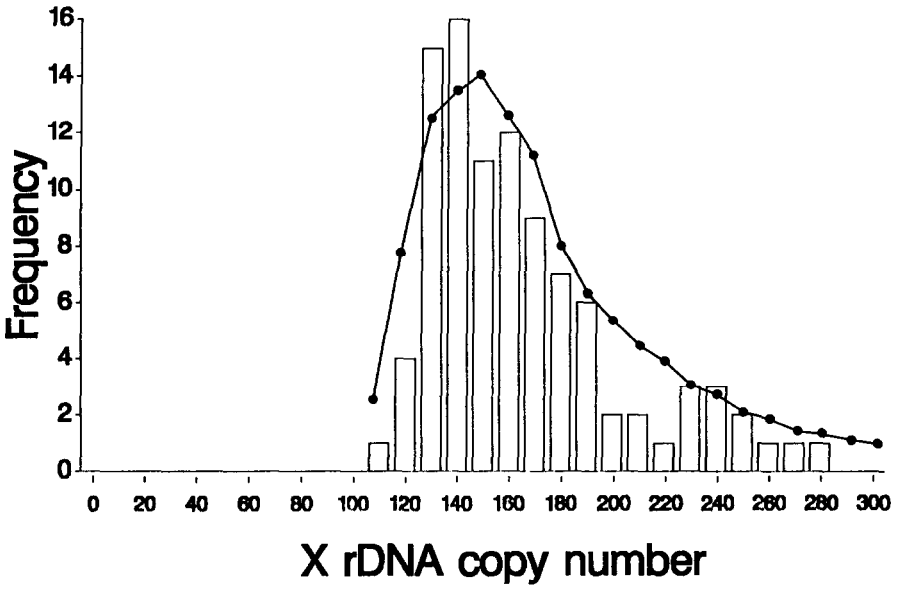

FIG. 2.—Observed X-chromosome rDNA copy-number distribution (bars) and minimum χ^2 fit of model (line). Parameters are from table 1, line 1.

variation at autosomal sites. If all lines have the same number of autosomal copies as do the two lines for which absolute copy numbers were estimated, then figure 4 reflects copy-number variation on the sex chromosomes only.

The copy-number distributions provide a snapshot of the status of the multigene

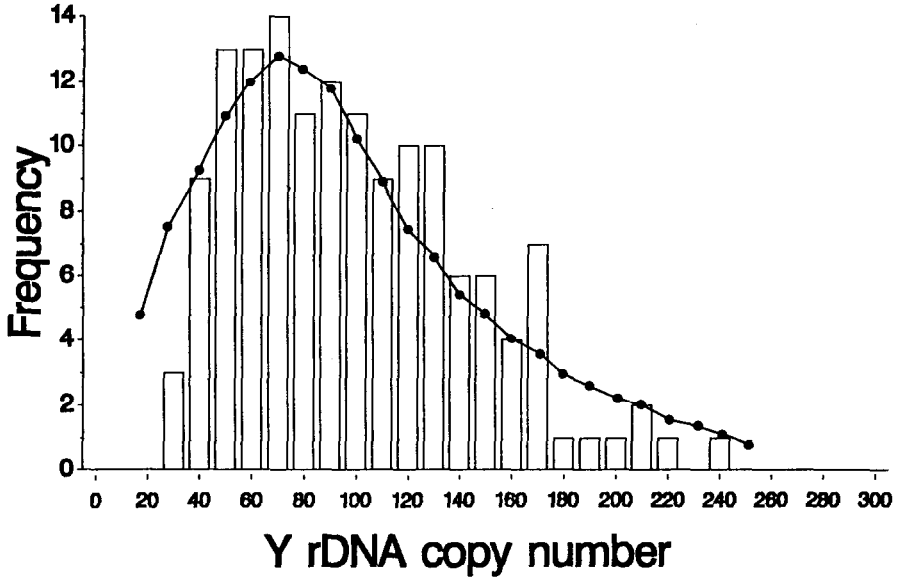

FIG. 3.—Observed Y-chromosome rDNA copy-number distribution (bars) and minimum χ^2 fit of model (line). Parameters are from table 1, line 6.

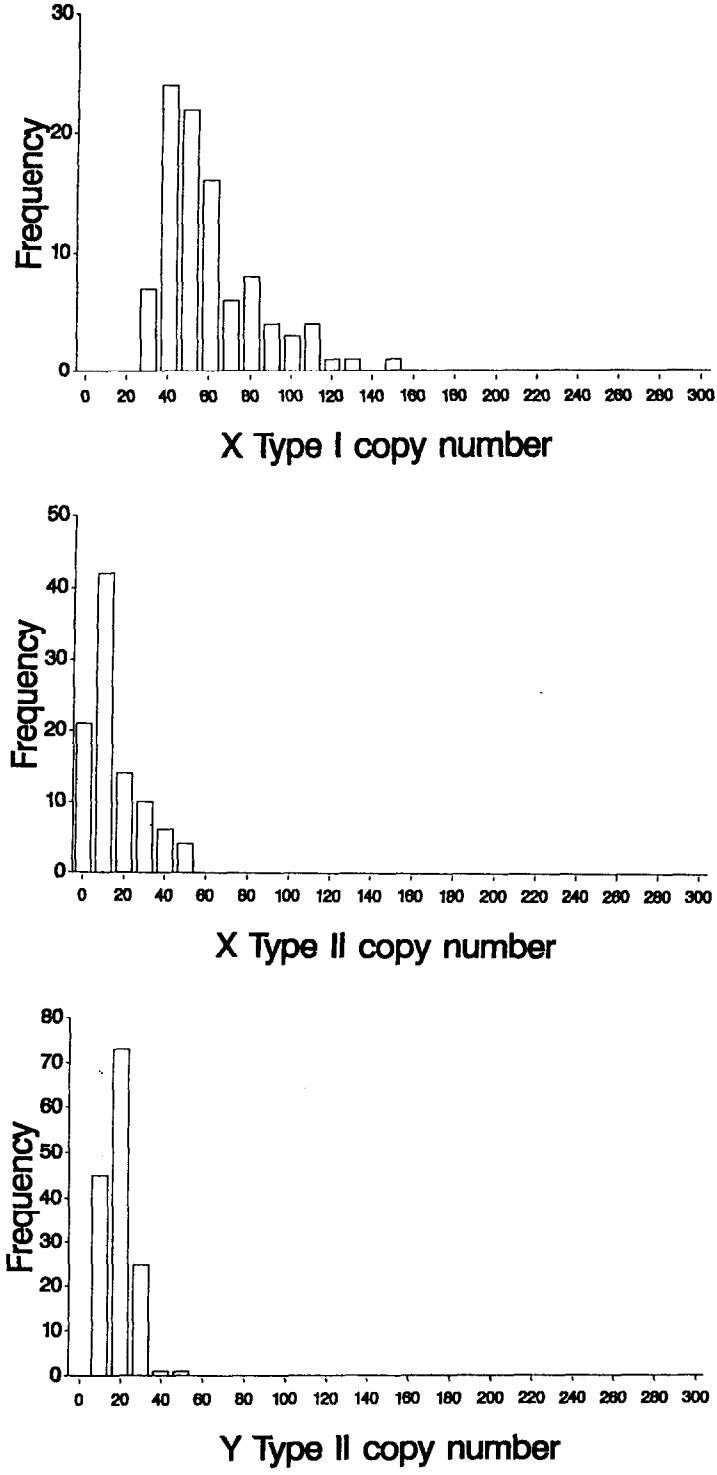

FIG. 4.—Copy-number distributions of type I inserts and type II inserts on X and Y chromosomes of *Drosophila melanogaster*.

family in the population. Among wild X and Y chromosomes of *Drosophila melanogaster,* there is a large variance in rDNA copy number, with the mean ± standard deviation being 165.0 ± 37.4 (range 113–275) on the X chromosome and 99.9 ± 44.2 (range 30–241) on the Y chromosome. A one-tailed F test on log-transformed data showed that this difference was significant ($F_{144,95} = 1.40, P < 0.01$). The distributions of rDNA copy number show a characteristically skewed shape, with a long tail to the right. The skewness of the distributions is particularly important in providing information about the rates of exchange events and about the operation of natural selection.

On the X chromosome there was a high correlation between the number of rDNA copies and the number of type I inserts (fig. 5). This implies that the proportion of X-linked copies of rDNA having a type I insert does not vary with rDNA copy number. This too is relevant to the models for the evolution of copy number, since type I insert–containing genes do not produce mature rRNA but are actively transcribed. The high correlation between rDNA copy number and type I insert copy number ($r = 0.86, P < 0.0001$) is consistent with the bulk of the variation in type I inserts residing in the X-linked 28S genes.

Model Fits

The best fits of models with various combinations of exchange events were obtained with a computer routine that used the simplex method to search the parameter space for the lowest χ^2 value. In the absence of stabilizing natural selection or exchange events that are biased toward lower-copy-number products, models with unequal crossover events tend to make the variance in copy number grow without bound. This is reflected in the χ^2 values in table 1, which show that models without intrachromatid

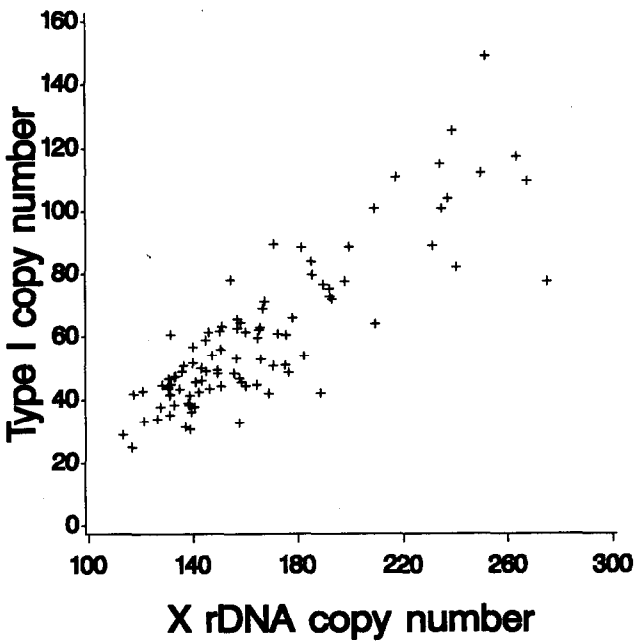

FIG. 5.—Scattergram of copy numbers of rDNA array on X chromosome, plotted against copy number of type I insert on X chromosome.

Table 1
Fits of Model to Observe *Drosophila* rDNA Copy-Number Distributions

	w_1	w_2	β	δ	γ	ω	χ^2	df
X only	50	140	0.0006	0.0142	0.0002	...	4.08	5
	...	...	0.0034	0.0040	0.0158	...	65.46	8
	100	140	...	0.0096	0.0001	...	8.28	6
	90	140	0.0006	...	0.0042	...	∞	
	90	140	0.0065	0.0042	...	...	7.96	6
Y only	20	70	0.0045	0.0227	...	...	12.29	7
	...	...	0.0096	0.0102	...	...	48.38	8
	20	60	...	0.0074	...	...	110.62	5
	20	70	0.0094	...	...	...	∞	
X and Y exchange allowed	20	70	0.0009	0.0184	0.0174	0.0003	27.46	12
	...	...	0.0205	0.0153	0.0	0.0091	191.30	13
	20	70	...	0.0110	0.0090	0.0006	29.78	13
	30	60	0.0017	...	0.0083	0.0097	∞	
	20	70	0.0005	0.0038	...	0.0004	27.72	13
	20	70	0.0004	0.0111	0.0091	...	28.42	13

NOTE.—w_1 is the copy number below which the flies cannot survive, and w_2 is the copy number above which the fitness is 1; the fitness of genotypes with intermediate copy numbers is a linear function between these truncation points. β is the rate of sister-chromatid exchange, δ is the rate of intrachromatid exchange, γ is the rate of interchromosomal unequal crossing-over between X and chromosomes, and ω is the rate of unequal crossing-over between the X- and Y-linked rDNA arrays; all four rates are expressed in terms of frequencies per chromosome. The χ^2 values indicate the goodness-of-fit between each model and the data. An ellipsis (. . .) indicates that parameter was fixed at 0. Absence of natural selection resulted in predicted distributions with too many low-copy-number chromosomes, and absence of intrachromatid exchange ($\delta = 0$) resulted in unbounded increase in χ^2.

exchange fit the data very poorly. Inclusion of intrachromatid exchange and natural selection against low-copy-number chromosomes results in reasonably good fits to the data. It is likely that models with stabilizing natural selection could fit the data, but because there is no evidence for a deleterious effect of high copy numbers, our objective was to see whether models without stabilizing selection could fit the data acceptably. In these models, fitness does not decrease with high copy number, and the rate of intrachromatid exchange must exceed the rate of sister-chromatid exchange. The low rate of X-Y interchromosomal exchange is suggested both by the separate fits to the X- and Y-chromosomal array with zero exchange and by the low values of ω from the model that treats both arrays jointly.

Discussion

The model-fitting approach suffers the problem that more than one model can fit the data adequately, so we cannot have complete confidence in the estimates of exchange rates. We saw, for example, that unequal sister-chromatid exchange and interchromosomal exchange had similar effects on the steady-state distribution of copy number. In particular, the model fits nearly as well with a high β and a low γ as it does with the reverse. Furthermore, the fact that a deterministic model fitted the data well does not imply that drift is an insignificant factor in the evolution of multigene families. With these caveats stated, we can emphasize some of the strong conclusions

that can be made from the data. The models require that natural selection eliminate the very-low-copy-number chromosomes from the population. No model that lacks natural selection can generate a steady-state distribution with the low density of low-copy-number chromosomes that was observed in nature. The fits of the models to the data also allow strong statements to be made about the relative rates of some exchange events. The model of Ohta and Dover (1983) showed that a multigene family that occurs on more than one chromosome will be nearly as homogenized as it is in the case of a single chromosome family, unless the rate of interchromosomal exchange is much lower than rates of intrachromatid exchange. The observations of differences between X- and Y-linked rDNA suggest that X-Y exchanges are much less frequent than intrachromatid exchanges, a conclusion that was strongly supported by the model fits. A like argument was made by Ohta (1990) in interpreting the patterns of variation among the human rDNA genes (Seperack et al. 1988) and the *Drosophila* histone genes (Matsuo and Yamazaki 1989).

The theory of copy-number evolution is directly coupled to the forces—including mutation, unequal homologous crossing-over, unequal sister-chromatid exchange, nonhomologous exchange, gene conversion, and random drift—that influence sequence divergence (Ohta 1980; Stephan 1989). Not only can the magnitudes of many of these forces be influenced by copy number, but the strength and nature of selection may be influenced as well, because tolerance of altered expression is likely to be greater with higher copy numbers. While unequal exchange events serve to homogenize sequences, they are also a source of variation in copy number. Although sister-chromatid exchange and interchromosomal exchange can have similar effects on the copy-number distribution, intra- and interchromatid gene conversion have very different consequences on the identities of genes on the same or different chromosomes (Ohta 1983; Ohta and Dover 1983). Drift and selection are the primary forces that can reduce the variance in copy number, so the observed variance in copy number is indicative of the relative rates of homogenization and generation of length variants. The occurrence of Y-chromosome polymorphism poses a problem in the context of deterministic population genetics, because models that exclusively invoke natural selection fail to maintain Y-linked polymorphism (Clark 1987). Clearly, drift and molecular-exchange events are essential to the maintenance of the polymorphism. The higher variance in Y-linked rRNA gene copy number appears contrary to the expectation based on a pure drift model, since the effective population size of the Y chromosome is smaller than that of the X chromosome (Charlesworth et al. 1987). An unresolved theoretical problem is the simultaneous maintenance of high levels of variability in both copy number and sequence, since an unequal exchange inflates copy-number variance but homogenizes the sequences of individual copies.

The model that is examined here, which is an extension of that of Takahata (1981) and Walsh (1987), does not lend itself well to analytical solution. Although Walsh (1987) showed that the distribution of copy number is bounded by a geometric distribution, our need to obtain a complete steady-state distribution of copy number required computer simulation. Our model differed from that of Takahata (who also used simulations) three ways: not allowing stabilizing selection, allowing intrachomatid exchange, and allowing X-Y exchange. The analysis of copy-number evolution of transposable elements has made use of multitype-branching-process models, which admit analytical solutions (Sawyer and Hartl 1986; Moody 1988). Dramatic differences between the results of these models and those for eukaryotes (Charlesworth and

Charlesworth 1983; Langley et al. 1983) underscore the importance that recombination has in the evolution of multiple-copy genes. The need to jointly consider copy number and identity evolution makes the most biologically interesting models intractable analytically (Ohta 1987, 1988). It would appear, then, that the future of the development of the theory of multigene-family evolution is relegated to computer simulation; and the purpose here was to show that there is some hope of parameterizing these models by fitting them to data.

Model fitting does not obviate the need for experimental attempts to obtain direct determinations of rates of exchange. Whether the rate of exchange depends on the number of gene copies has been a focus of many experiments on the rDNA array in *D. melanogaster*. These experiments have made use of the fact that low-copy-number flies have the *bb* phenotype, making it possible to infer changes in gene multiplicity by scoring this phenotype. The disadvantage to this approach is that most of our knowledge of copy-number changes in the rDNA array comes from the study of chromosomes with abnormally low copy numbers. Bobbed males, which have a low copy number on both the X and Y chromosomes, can produce ≤90% revertants among their progeny, in a process known as magnification (Ritossa 1968; Tartof 1974*a*, 1974*b*). These reversions occur in clusters, implying that they involve premeiotic events. The failure to observe *bobbed* progeny from flies with normal copy number suggests that the array is very stable when sufficient copies are present. The apparent tendency for only low-copy-number chromosomes to magnify has important implications for the evolution of rRNA gene copy number. An important component of this copy-number dependence is that cell-lineage selection (clonal amplification) may occur, favoring those pregerm cells that have an increase in copy number over those that are deficient. This cellular selection would result in more magnified gametes being produced by *bobbed* flies, even if the rate of the exchange events is independent of copy number. In experiments where cellular selection can be controlled, the rates of exchange do appear to be independent of copy number (Endow et al. 1984; Hawley and Marcus 1989). The failure of ring X chromosomes to magnify provides clear evidence for the involvement of unequal sister-chromatid exchange. Because such exchanges result in unstable X chromosomes (Endow et al. 1984; Endow and Komma 1986).

There are a number of diagnostic differences between the X- and Y- linked rDNA arrays. On average, the Y-linked rDNAs have longer intergenic sequences than do X-linked copies (Coen and Dover 1983), and more than half of the X repeats have insertions that are rare on the Y (Wellauer et al. 1978; de Cicco and Glover 1983). Furthermore, despite the extensive similarities in restriction maps, there are diagnostic sequence differences in the 18S gene (Yagura et al. 1979). A survey of restriction-site variation revealed that the Y chromosome had greater interpopulation variation in sequence than did X-linked sequences (Williams et al. 1987). These observations clearly show that the rate of intrachromosomal exchange is greater than the rate of interchromosomal exchange. Experiments that directly recover unequal exchange events yield estimates in the range of 10^{-4}–10^{-5} for the rate of unequal crossing-over within rDNA (Maddern 1981; Williams et al. 1989). The X-Y exchanges were also clustered among the progeny, so these events are premeiotic. Response to artificial selection for high and low abdominal bristle counts was shown to be at least partially mediated by the rDNA array (Frankham et al. 1980). Response was greater in females than in males, corresponding to observed changes in X-linked but not in Y-linked

rDNA copy number. Selection experiments yielded an estimated rate of unequal exchange of 3×10^{-4}/gamete generation, and a definitive X-Y exchange product was recovered in the form of a compound X-Y translocation (Coen and Dover 1983). Rates of unequal exchange are apparently great enough to result in concerted evolution, thereby homogenizing sequences on a chromosome (Coen et al. 1982), and this might be expected to generate and maintain copy-number variation as well.

With the exception of intrachromatid exchange, the exchange events depicted in figure 1 are expected to result in an equal number of gametes having increases and decreases in copy number. The cell-lineage selection mentioned above results in an excess of high-copy-number gametes, and, under magnifying conditions, the frequency of reductions in copy number is $\sim 1/10$ that of increases in copy number. Evidence for cell-lineage selection comes from the observation of clustering—high-copy-number gametes occur in clusters because of selection among premeiotic cells, but reductions, which are not selected, do not occur in clusters (Tartof 1974*a*; Hawley and Marcus 1989). When a correction is made for cell-lineage selection, the bias toward increased copy number disappears (Hawley and Tartof 1985). The model that was fitted in the present report did not include cell-lineage selection, but the way it would be modeled is very close to the way that natural selection was. We assumed that fitness was a function of the sum of the copy numbers on the two homologous chromosomes and that this additive form of selection would not differ from the gametic cell-lineage selection. The important point is that, by allowing selection to occur at the cellular level, one could obtain fits between a model with no natural selection and the observed copy-number distribution.

Because we do not know all of the biological mechanisms involved in the regulation of rDNA copy number, it may be thought premature to attempt to model the phenomenon. Our objective was to ask whether the known types of exchange events could produce a copy-number distribution that was similar to the observed distribution, and we obtained a clearly affirmative answer. We are careful to point out that this does not provide conclusive evidence that this is the only possible mechanism for copy-number regulation. One aspect of the rDNA array that was not included in the model was the occurrence of type I and type II inserts. These inserts show sequence similarity to non-LTR retrotransposons, and, in *Bombyx mori,* products of similar elements exhibit sequence-specific endonuclease activity (Xiong and Eickbush 1988). The observation that type II inserts are amplified 16-fold more in males undergoing rDNA magnification than in nonmagnifying female controls (Labella et al. 1983) led Hawley and Marcus (1989) to propose that this overexpression at low copy number results in increased frequency of endonuclease-induced nicks, which provides the opportunity for unequal exchange events. This model is highly speculative, but, if it is correct that the type II inserts play such a key role in magnification, then this fact would help explain the persistence of these elements. The fact that the elements can transpose (Jakubczak et al. 1990) adds another intriguing complication to the evolution of the rDNA array. Another potentially significant attribute of rDNA was not considered in the present paper—i.e., the fact that the rDNA array serves as the pairing site for the sex chromosomes in male meiosis (McKee and Karpen 1990). Constraints on pairing may be another function of the rDNA array relevant to the regulation of copy number. Both incorporation of type I and type II insert transposition and consideration of pairing will be considered in models presented elsewhere.

Acknowledgments

We thank Sharyn Endow, Stephen Schaeffer, and Scott Williams for sharing plasmid clones, and we thank Karen Bell, Sharyn Houtz, and Frances Szumski for providing technical assistance. Sharyn Endow and anonymous reviewers provided helpful comments on the manuscript. This work was supported by NIH grants HD00743 and HD21963.

LITERATURE CITED

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

CHARLESWORTH, B., J. A. COYNE, and N. H. BARTON. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130**:113–146.

CLARK, A. G. 1987. Natural selection and Y-linked polymorphism. Genetics **115**:569–577.

CLARK, A. G., and E. M. S. LYCKEGAARD. 1988. Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mtDNA in *Drosophila melanogaster.* Genetics **118**:471–481.

COEN, E. S., and G. A. DOVER. 1983. Unequal exchanges and the coevolution of X and Y rDNA arrays in *Drosophila melanogaster.* Cell **33**:849–855.

COEN, E. S., J. M. THODAY, and G. A. DOVER. 1982. Rate of turnover of the structural variants in the rDNA gene of *Drosophila melanogaster.* Nature **295**:564–568.

CROW, J. F., and M. KIMURA. 1970. An introduction to population genetics theory. Harper & Row, New York.

DAWID, I. B., P. K. WELLAUER, and E. O. LONG. 1978. Ribosomal DNA in *Drosophila melanogaster.* I. Isolation and characterization of cloned fragments. J. Mol. Biol. **126**:749–768.

DE CICCO, D. V., and D. M. GLOVER. 1983. Amplification of rDNA and type I sequences in *Drosophila* males deficient in rDNA. Cell **32**:1217–1225.

ENDOW, S. A. 1982. Polytenization of the ribosomal genes on the X and Y chromosomes of *Drosophila melanogaster.* Genetics **100**:375–385.

ENDOW, S. A., and D. J. KOMMA. 1986. One-step and stepwise magnification of a bobbed lethal chromosome in *Drosophila melanogaster.* Genetics **114**:511–523.

ENDOW, S. A., D. J. KOMMA, and K. C. ATWOOD. 1984. Ring chromosomes and rDNA magnification in *Drosophila.* Genetics **108**:969–983.

FRANKHAM, R., D. A. BRISCOE, and R. K. NURTHEN. 1980. Unequal crossing over at the rRNA tandon as a source of quantitative genetic variation in *Drosophila.* Genetics **95**:727–742.

GLOVER, D. M., and D. S. HOGNESS. 1977. A novel arrangement of the 18S and 28S sequences in a repeating unit of *Drosophila melanogaster* rDNA. Cell **10**:167–176.

GOLDBERG, D. A. 1980. Isolation and partial characterization of the *Drosophila* school dehydrogenase gene. Proc. Natl. Acad. Sci. USA **77**:5794–5798.

HAWLEY, R. S., and C. R. MARCUS. 1989. Recombinational controls of rDNA redundancy in *Drosophila.* Annu. Rev. Genet. **23**:87–120.

HAWLEY, R. S., and K. D. TARTOF. 1985. A two-stage model for the control of rDNA magnification. Genetics **109**:691–700.

JAKUBCZAK, J. L., Y. XIONG, and T. H. EICKBUSH. 1990. Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila* are retrotransposable elements closely related to those of *Bombyx mori.* J. Mol. Biol. **212**:37–52.

KIDD, S. J., and D. M. GLOVER. 1981. *Drosophila melanogaster* ribosomal DNA containing type II insertions is variably transcribed in different strains and tissues. J. Mol. Biol. **151**: 645–662.

KOMMA, D. J., and S. A. ENDOW. 1987. Incomplete Y chromosomes promote magnification in male and female *Drosophila.* Proc. Natl. Acad. Sci. USA **84**:2382–2386.

LABELLA, T., L. VICARI, A. MANZI, and F. GRAZIANA. 1983. Expression of rDNA insertions during rDNA magnification in *Drosophila melanogaster.* Mol. Gen. Genet. **190**:487–493.

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

LONG, E. O., M. COLLINS, B. I. KIEFER, and I. B. DAWID. 1981. Expression of the ribosomal DNA insertions in *bobbed* mutants of *Drosophila melanogaster.* Mol. Gen. Genet. **182**:377–384.

LONG, E. O., and I. B. DAWID. 1979. Expression of ribosomal DNA insertions in *Drosophila melanogaster.* Cell **18**:1185–1196.

———. 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. **49**:727–764.

LONG, E. O., M. L. REBBERT, and I. B. DAWID. 1980. Structure and expression of ribosomal RNA genes of *Drosophila melanogaster* interrupted by type-2 insertions. Cold Spring Harbor Symp. Quant. Biol. **45**:667–672.

LYCKEGAARD, E. M. S., and A. G. CLARK. 1989. Ribosomal DNA and *Stellate* gene copy number variation on the Y chromosome of *D. melanogaster.* Proc. Natl. Acad. Sci. USA **86**:1944–1948.

MCKEE, B. D., and G. H. KARPEN. 1990. *Drosophila* ribosomal RNA genes function as an X-Y pairing site during male meiosis. Cell **61**:61–72.

MADDERN, R. H. 1981. Exchange between the ribosomal RNA genes of X and Y chromosomes of *Drosophila melanogaster* males. Genet. Res. Camb. **38**:1–7.

MATSUO, Y., and T. YAMAZAKI. 1989. Nucleotide variation and divergence in the histone multigene family of *Drosophila melanogaster.* Genetics **122**:87–97.

MOODY, M. E. 1988. A branching process model for the evolution of transposable elements. J. Math. Biol. **26**:347–357.

OHTA, T. 1980. Evolution and variation of multigene families. Springer, Berlin.

———. 1983. Theoretical study on the accumulation of selfish DNA. Genet. Res. **41**:1–15.

———. 1987. Simulating evolution by gene duplication. Genetics **115**:207–213.

———. 1988. Further simulation studies on evolution by gene duplication. Evolution **42**:375–386.

———. 1990. How gene families evolve. Theor. Popul. Biol. **36**:213–219.

OHTA, T., and G. A. DOVER. 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. Proc. Natl. Acad. Sci. USA **80**:4079–4083.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY, and W. T. VETTERLING. 1986. Downhill simplex methods in multidimensions. Pp. 289–293 *in* Numerical recipes—the art of scientific computing. Cambridge University Press, Cambridge.

RITOSSA, F. 1968. Unstable redundancy of genes for ribosomal RNA. Proc. Natl. Acad. Sci. USA **60**:509–516.

———. 1976. The *bobbed* locus. Pp. 801–846 *in* M. Ashburner and E. Novitski, eds. The genetics and biology of *Drosophila.* Vol. **1b**. Academic Press, New York.

SAWYER, S. A., and D. L. HARTL. 1986. Distribution of transposable elements in prokaryotes. Theor. Popul. Biol. **30**:1–17.

SEPERACK, P., M. SLATKIN, and N. ARNHEIM. 1988. Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. Genetics **119**:943–949.

STEPHAN, W. 1989. Tandem-repetitive noncoding DNA: forms and forces. Mol. Biol. Evol. **6**: 198–212.

TAKAHATA, N. 1981. A mathematical study on the distribution of the number of repeated genes per chromosome. Genet. Res. Camb. **38**:97–102.

TARTOF, K. D. 1971. Increasing the multiplicity of ribosomal RNA genes in *Drosophila melanogaster.* Science **171**:294–297.

TARTOF, K. D. 1974*a*. Unequal mitotic sister chromatid exchange as the mechanism of ribosomal gene magnification. Proc. Natl. Acad. Sci. USA **71**:1272–1276.

———. 1974*b*. Unequal sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. Cold Spring Harbor Symp. Quant. Biol. **38**:491–500.

TAUTZ, D., J. M. HANCOCK, D. A. WEBB, C. TAUTZ, and G. A. DOVER. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster.* Mol. Biol. Evol. **5**:366–376.

TERRACOL, R., and N. PRUD'HOMME. 1986. Differential elimination of *bobbed* mutants of *Drosophila melanogaster.* Mol. Cell. Biol. **6**:1023–1031.

WALSH, J. B. 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics **115**:553–567.

WELLAUER, P. K., I. B. DAWID, and K. D. TARTOF. 1978. X and Y chromosomal ribosomal DNA of *Drosophila:* comparison of spacers and insertions. Cell **14**:269–278.

WILLIAMS, S. M., G. R. FURNIER, E. FUOG, and C. STROBECK. 1987. Evolution of the ribosomal DNA spacers of *Drosophila melanogaster:* different patterns of variation on X and Y chromosomes. Genetics **116**:225–232.

WILLIAMS, S. M., J. A. KENNISON, L. G. ROBBINS, and C. STROBECK. 1989. Reciprocal recombination and the evolution of the ribosomal gene family of *Drosophila melanogaster.* Genetics **122**:617–624.

XIONG, Y., and T. H. EICKBUSH. 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. **8**: 114–123.

YAGURA, T., M. YAGURA, and M. MURAMATSU. 1979. *Drosophila melanogaster* has different ribosomal RNA sequences on X and Y chromosome. J. Mol. Biol. **133**:533–547.

BRIAN CHARLESWORTH, reviewing editor

Received August 20, 1990; revision received December 18, 1990

Accepted January 9, 1991