The Molecular Through Ecological Genetics of *abnormal abdomen*. IV. Components of Genetic Variation in a Natural Population of *Drosophila mercatorum*

Hope Hollocher,* Alan R. Templeton,* Rob DeSalle† and J. Spencer Johnston‡

**Department of Biology, Washington University, St. Louis, Missouri 63130, †Department of Biology, Yale University, New Haven, Connecticut 06511, and ‡Department of Entomology, Texas A&M University, College Station, Texas 77843*

Manuscript received September 24, 1990
Accepted for publication October 15, 1991

ABSTRACT

Natural populations of *Drosophila mercatorum* are polymorphic for a phenotypic syndrome known as *abnormal abdomen* (*aa*). This syndrome is characterized by a slow-down in egg-to-adult developmental time, retention of juvenile abdominal cuticle in the adult, increased early female fecundity, and decreased adult longevity. Previous studies revealed that the expression of this syndrome in females is controlled by two closely linked *X* chromosomal elements: the occurrence of an R1 insert in a third or more of the *X*-linked 28S ribosomal genes (rDNA), and the failure of replicative selection favoring uninserted 28S genes in larval polytene tissues. The expression of this syndrome in males in a laboratory stock was associated with the deletion of the rDNA normally found on the *Y* chromosome. In this paper we quantify the levels of genetic variation for these three components in a natural population of *Drosophila mercatorum* found near Kamuela, Hawaii. Extensive variation is found in the natural population for both of the *X*-linked components. Moreover, there is a significant association between variation in the proportion of R1 inserted 28S genes with allelic variation at the underreplication (*ur*) locus such that both of the necessary components for *aa* expression in females tend to cosegregate in the natural population. Accordingly, these two closely linked *X* chromosomal elements are behaving as a supergene in the natural population. Because of this association, we do not believe the R1 insert to be actively transposing to an appreciable extent. The *Y* chromosomes extracted from nature are also polymorphic, with 16% of the *Y*s lacking the *Y*-specific rDNA marker. The absence of this marker is significantly associated with the expression of *aa* in males. Hence, all three of the major genetic determinants of the *abnormal abdomen* syndrome are polymorphic in this natural population.

IN 1975, a polymorphic syndrome known as *abnormal abdomen* (*aa*) was discovered in natural populations of *Drosophila mercatorum* living near Kamuela, Hawaii. The syndrome was first identified because it slowed down the egg-to-adult developmental time of females, but many other pleiotropic effects were soon discovered; including the retention of juvenile abdominal cuticle in the adult, increased early fecundity in adult females, and decreased adult longevity (TEMPLETON and RANKIN 1978; TEMPLETON 1982, 1983). A Mendelian genetic analysis revealed that the *aa* syndrome depends upon two *X*-linked elements that are about half a map unit apart on the heterochromatic end of the *X* chromosome (TEMPLETON, CREASE and SHAH 1985). In addition, minor modifiers of the morphological expression of *aa* can be found on all the autosomes, and a *Y*-linked element controls the morphological expression in males (TEMPLETON, CREASE and SHAH 1985).

Some of the phenotypic effects of *aa* (the developmental slowdown and juvenilized abdominal cuticle) are similar to *bobbed* phenotypes in *Drosophila melanogaster* and *Drosophila hydei*, which are known to be caused by a deficiency in the number of functional 18S/28S ribosomal genes (RITOSSA 1976; FRANZ and KUNZ 1981; TERRACOL and PRUD'HOMME 1986). Moreover, the 18S/28S rDNA cluster is known to be located at the heterochromatic end of the X chromosome and on the *Y* chromosome, the same genetic locations for the major determinants of *aa*. Accordingly, rDNA molecular studies were initiated. DESALLE, SLIGHTOM and ZIMMER (1986) discovered that there is no deficiency in the amount of rDNA in lines displaying the morphological effects of *aa*, but that these *aa* lines did have a high proportion of their 28S genes bearing a 5kb insert in the β portion of the coding region. The inserted 28S genes are clustered together, and few if any copies of the insert are found outside the *X*-linked rDNA (DESALLE, SLIGHTOM and ZIMMER 1986).

D. mercatorum is not unique in having interrupted 28S genes. The discovery that the 28S rRNA genes of Drosophila often carry insertion sequences not coding for rRNA was first discovered in the late 1970s (WHITE and HOGNESS 1977; WELLAUER and DAWID 1977, 1978; DAWID, WELLAUER and LONG 1978;

WELLAUER, DAWID and TARTOF 1978; ROIHA and GLOVER 1980). The insertion sequences were classified as being either type I or type II (now referred to as R1 or R2) depending on the specific insertion site in the 28S gene. The insert type representing the 5-kb insert of *D. mercatorum* has been identified as being of the R1 variety (T. H. EICKBUSH, personal communication). The phenomenon of having interrupted rRNA gene sequences has since been found in other invertebrates as well, such as *Calliphora erythrocephala* (BECKINGHAM and WHITE 1980), *Bombyx mori* (LECANIDOU, EICKBUSH and KAFATOS 1984), *Ascaris lumbricoides* (BACK *et al.* 1984), two species of *Anopheles* (PASKEWITZ and COLLINS 1989), and, more recently, a whole range of insect species covering nine orders (JAKUBCZAK, BURKE and EICKBUSH 1991). Detailed analysis of the R1 and R2 elements of Drosophila and of Bombyx indicates these insertions may be retrotransposons based on their genetic structure and sequence similarity to the reverse transcriptase gene, and to the *gag* gene in the case of R1 (EICKBUSH and ROBINS 1985; BURKE, CALALANG and EICKBUSH 1987; XIONG *et al.* 1988; XIONG and EICKBUSH 1988a; JAKUBCZAK, XIONG and EICKBUSH 1990). The evidence that these retrotransposons may still be active comes from a positive *in vitro* assay of site-specific endonuclease activity encoded by the R2 element that may be responsible for the site specific insertion of these elements (XIONG and EICKBUSH, 1988b), although direct evidence of transposition activity has not been documented.

Repeats that contain an insert are transcribed inefficiently (LONG and DAWID 1979; LONG, REBBERT and DAWID 1980; LONG *et al.* 1981; TERRACOL 1986), or, when under strict circumstances they are transcribed, they may not be processed properly (MAKNI, MARRAKCHI and PRUD'HOMME 1989). Hence, 18S/28S units bearing the R1 insert are functionally inactivated, and *X* chromosomes bearing a large proportion of inserted 28S genes have a deficiency of functional rDNA even though the total amount of rDNA may be normal. By screening a large number of *aa* stocks, we have found that all stocks with morphological expression have at least a third or more of the 28S rDNA genes bearing the R1 insert (TEMPLETON *et al.* 1989).

The above condition appears necessary, but not sufficient, to insure morphological expression of the *aa* syndrome. DESALLE and TEMPLETON (1986) discovered that a second criterion must also be satisfied: there must be no preferential underreplication (replicative selection in the terminology of GOODRICH-YOUNG and KRIDER 1989) of inserted rDNA repeats in polytene tissues such as the larval fat body. Euchromatic DNA is greatly amplified in polytene tissues, but the rDNA tends to be underreplicated relative to the euchromatin [reviewed in BECKINGHAM (1982), KORGE (1987), and SRADLING and ORR-WEAVER (1987)]. It has been well documented in *D. melanogaster* that selective replication can favor certain repeat types within the nucleolus organizer [reviewed in KORGE (1987) and SPRADLING (1987)], and the same is true for *D. mercatorum*. In *aa* flies, the underreplication is uniform across the rDNA cluster, but in flies that do not express *aa* but nevertheless have a large portion of their 28S genes bearing the R1 insert, there is preferential underreplication of the inserted 28S repeats (DESALLE and TEMPLETON 1986). Because of this preferential underreplication, the uninserted functional 28S repeats are effectively overreplicated relative to the nonfunctional inserted repeats. The presence or absence of preferential underreplication is controlled by an *X*-linked locus (the underreplication or *ur* locus), with the allele coding for no preferential underreplication acting as a recessive for the morphological effects of the *aa* syndrome. The dominant allele causes preferential underreplication for both *cis* and *trans* rDNA repeats. The *Y* chromosomal rDNA can also be effectively overreplicated in males, thereby suppressing the syndrome (DESALLE and TEMPLETON 1986). A spontaneous deletion of the *Y*-linked rDNA allows the expression of *aa* in males (DESALLE, SLIGHTOM and ZIMMER 1986).

In summary, these studies involving laboratory stocks indicate that the *aa* syndrome in *D. mercatorum* females requires two *X*-linked molecular events: (1) about a third or more of the 28S rDNA repeats must bear an R1 insert, and (2) there must be no preferential underreplication of inserted 28S genes in the larval fat body. The morphological expression of *aa* in males in addition requires the deletion of the rDNA found on the *Y* chromosome. The purpose of this paper is to quantify the genetic variation found for these two *X*-linked and one *Y*-linked components of the *aa* syndrome in a natural population of *D. mercatorum* near Kamuela, Hawaii.

MATERIALS AND METHODS

Drosophila collections, stocks and culture conditions: To test for the morphological expression of *aa*, we use crosses to a laboratory *aa* strain, as will be discussed subsequently. This *aa* strain was derived from the Kamuela, Hawaii, natural population through an artificial selection scheme to enhance penetrance and expression (TEMPLETON and RANKIN 1978). A spontaneous deletion of the *Y*-linked rDNA allowed expression of *aa* in males (DESALLE, SLIGHTOM and ZIMMER 1986), and the *aa* stock used in these experiments bears that original deletion. In order to keep the penetrance and extent of morphological expression high, the stock is subjected to artificial selection for juvenilized cuticle on a periodic basis. The current penetrance is nearly 100% in both males and females, with over 90% of both sexes having over 50% of their abdominal cuticle juvenilized. The stock S-1-Br16 (originally derived from a

collection made in El Salvador) was used as a wild-type standard.

All collections of the natural populations took place near Kamuela, Hawaii. Near and around this town are extensive patches of the prickly pear cactus, *Opuntia megacantha*, which serves as the sole larval food resource for the Hawaiian populations of *D. mercatorum* (JOHNSTON and TEMPLETON 1982). Flies were captured by direct aspiration off of rotting cladodes and aspiration of flies attracted to rotten guava placed inside gallon-sized zip lock bags that in turn are placed well inside a cactus patch. The collections used in this study were made in July-August 1981, June 1982, August 1983, June-July 1984, December 1986, and December 1987-January 1988. As will be described shortly, wild-caught females were used to establish isofemale lines, and their male offspring were then crossed to the *aa* stock. Wild-caught males were mated directly to *aa* females. Only males were used in these assays in order to take advantage of a *Y*-specific rDNA marker that allows us to partition the molecular genetics of the *aa* system into *X* and *Y* chromosome components. There are no rDNA deletion stocks available for *D. mercatorum*; therefore, to assay individual *X* and *Y* chromosomes from nature for rDNA, we relied on the shorter nontranscribed spacer of the *Y* chromosome to distinguished *Y* rDNA repeats from *X* rDNA repeats, which were then further partitioned into insert and non-insert-bearing repeat types.

Stocks were maintained on standard cornmeal, agar, and corn syrup medium. All crosses involved in scoring for *aa* expression were performed at 25°. Laboratory stocks were maintained in pint bottles, and isofemale lines and all crosses of sons or wild-caught males to *aa* females were maintained in shell vials.

DNA extractions and Southern blot analysis: DNA from single adult male flies was isolated by the methods described in DESALLE, SLIGHTOM and ZIMMER (1986) and DESALLE and TEMPLETON (1986). The extracted DNA was digested with the restriction enzyme *Eco*RI, as directed by the supplier (New England Biolabs). The cut DNA was fractionated on 0.8% agarose gels, transferred to nylon or nitrocellulose filter paper, and hybridized with ^{32}P-labeled probe, pDmrY22 (DAWID, WELLAUER and LONG 1978), a full length uninterrupted rDNA repeat from *D. melanogaster* (kindly supplied by IGOR DAWID). The filters were washed and exposed to preflashed X-ray film as described in MANIATIS, FRITSCH and SAMBROOK (1982).

Determination of the proportion of inserted and non-inserted *X*-linked rDNA repeats, and the presence of *Y*-linked rDNA repeats: After being crossed to the *aa* stock, *D. mercatorum* males were frozen at −80° and subjected at some later time to the Southern blot analysis described above. The resulting autoradiographs were scanned with a densitometer. TEMPLETON *et al.* (1989) describe the results of preliminary experiments that show that (1) band areas and peak heights are not significantly different from one another in estimating the relative proportions of bands on the autoradiograph, (2) the measurement error associated with rescanning the same individuals is minor and not statistically significant whereas the differences between individuals are large and highly significant, and (3) different densitometers (ordinary light *vs.* laser) give virtually identical results that are no greater than the measurement error between replicates on the same densitometer. Because of these desirable measurement properties, we use peak heights on the densitometer scans as the raw data from which to estimate R1 insert proportions.

Figure 1 gives an example of an autoradiograph along with the restriction maps. As can be seen from the maps,

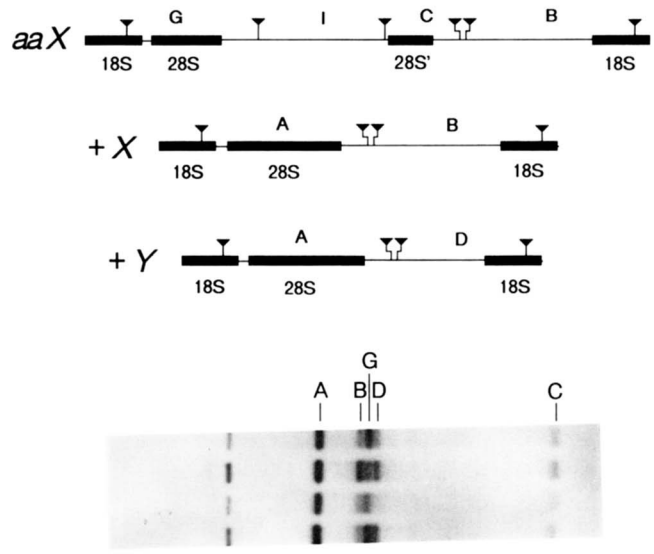

FIGURE 1.—Restriction maps of the major types of 18S/28S rDNA repeats found on the *X* and *Y* chromosomes of *D. mercatorum*, along with a Southern blot of *Eco*RI-digested genomic DNA of a typical male hybridized to a *D. melanogaster* 18S/28S rDNA probe. Bands are designated by the letters A, B, C, D, G and I.

two bands are specific to inserted rDNA repeats (G and C), one band is specific to noninserted repeats on both the *X* and *Y* chromosomes (A), one band is specific to *X* chromosomes for both inserted and noninserted repeats (B), and one band is specific to *Y* chromosomes and is found only in males (D). Some B bands may be associated with *Y*-linked rDNA, although this is not true for the laboratory lines we have characterized. Nevertheless, wild-extracted *Y*s have not been examined in detail to eliminate this possibility. In the present study it is assumed that all *Y*-linked rDNA repeats have the spacer length associated with the D band in Figure 1. Peak heights from these various bands can be combined in several alternative ways to estimate the proportion of the *X*-linked rDNA repeats that bear the R1 insert. We have chosen the following estimate (see the APPENDIX for discussion of the statistical attributes of this estimator):

$$G/\{((B - C)/(B - C + D))(A + G)\}. \quad (1)$$

Repeatability analysis of this estimate for a set of 33 individuals digested and run two separate times gave an R^2 of 0.9698 indicating that 97% of the variation could be accounted for by between individual differences rather than differences between replicates. A *t*-test analysis of the same data also showed that the replicate estimates of the R1 insert proportion for an individual were not significantly different on separate gels ($t = -0.3544$, NS, d.f. = 32). Therefore, the measure we are using is highly consistent and capable of ascribing different genotypes to the individuals in our assay.

As shown in Figure 1, the rDNA repeats found on the *Y* chromosome are associated with the presence of the D band upon Southern analysis. Deletion of this D band in laboratory stocks allows expression of the *aa* syndrome in males (DESALLE, SLIGHTOM and ZIMMER 1986). We therefore scored the autoradiographs from the wild-caught males and sons of wild-caught females for the presence or absence of the D band.

Determination of the genetic state at the underreplication locus: It is impractical to do large-scale population screening for the underreplication polymorphism using the procedures given in DESALLE and TEMPLETON (1986),

which involved the isolation of DNA from the larval brains and fat bodies of the same individuals. Accordingly, an alternative procedure for population screening was developed. A standard *aa* stock had been bred which has extremely high penetrance of juvenilized cuticle (TEMPLETON, CREASE and SHAH 1985). Because males always donate their *X* chromosome to their daughters and their *Y* chromosome to their sons, the genetic state of the *X* and *Y* chromosomes of males from nature can be assayed by scoring the resulting female and male progeny from a testcross to the standard *aa* stock. As shown in the APPENDIX, this *aa* stock has 76% of its 28S genes bearing the R1 insert and is homozygous for the recessive allele leading to no preferential underreplication. When a wild-caught male (or son of a wild-caught female) is crossed to an *aa* female, the resulting female offspring will have at least 38% of their 28S genes bearing the R1 insert even if the *X* chromosome from the wild-caught male bears no inserts at all. Hence, the resulting female progeny should always have sufficient inserted 28S genes to support the morphological expression of *aa* regardless of the insert proportion of the *X* chromosome extracted from nature. This expression will only occur if these females are homozygous for the allele causing no preferential underreplication. Consequently, if the female testcross progeny display juvenilized abdominal cuticle, the *X* chromosome extracted from the natural population must bear the allele for no preferential underreplication. We score 50 female progeny, and only make the inference of *abnormal abdomen* if two or more progeny have juvenilized cuticle. This criterion was chosen because *D. mercatorum* females have a very low parthenogenetic capacity (TEMPLETON 1979), so it is possible for even a mated female to produce an occasional parthenogenetic daughter, which would be *aa*. Moreover, the penetrance of *aa* is often incomplete, so it is essential to look at several flies. By scoring 50 daughters from each cross, we also obtain quantitative estimates on penetrance (the percentage of daughters with juvenilized cuticle) and expressivity (the average amount of abdominal cuticle that is juvenilized given expression). Expressivity is quantified on a 1 to 3 scale as follows: 1, only a small patch of abdominal cuticle is juvenilized; 2, two or more small patches or one large patch covering at least 5% but less than 50% of the abdominal cuticle is juvenilized; 3, 50% or more of the abdominal cuticle is juvenilized.

To see if this testing scheme works, we crossed the stock K28-0-Im to the *aa* stock. This stock had been thoroughly characterized in our molecular studies (DESALLE, SLIGHTOM and ZIMMER 1986; DESALLE and TEMPLETON 1986). The K28 *X* chromosome has very few inserted 28S genes (DESALLE, SLIGHTOM and ZIMMER 1986), yet this chromosome also codes for no preferential underreplication (DESALLE and TEMPLETON 1986). Of the female offspring from this cross, 64% had juvenilized cuticle. This shows that the underreplication polymorphism can be scored with this testcross procedure even when the input *X* chromosome has very few inserts.

Male progeny from the testcrosses to *aa* females described above were also scored for *aa* expression. These sons would receive the *aa X* chromosome from the laboratory stock, but their *Y* chromosome would be derived from nature. Since the *X* chromosome has 76% of its rDNA repeats bearing the R1 insert and because the sons are hemizygous for the allele leading to no preferential underreplication, these sons should display juvenilized cuticle unless their wild-derived *Y* chromosome can suppress the expression of *aa*. The sons were scored for penetrance and expressivity in exactly the same manner described above for their sisters.

To determine the concordance of *aa* penetrance in male *vs.* female testcross progeny, we used the same testcross criteria given above, and separately scored for *aa* in the sons and daughters of 268 wild-caught females crossed to the *aa* stock. In our wild-caught collections, we frequently found females that were either uninseminated or had exhausted their sperm load by the time they are taken back into the laboratory. These wild-caught females were mated to males from the *aa* stock. Hence, all the male offspring from these matings bear a *Y* chromosome that lacks rDNA. The daughters from these crosses receive one *X* chromosome from the *aa* tester stock, and the other from their wild-caught mother. If one or both of the *X* chromosomes borne by the mother is *aa*, both the daughters and the sons should show *aa* expression. If *aa* were completely penetrant in both sexes, we would expect complete concordance for sons and daughters, but if this syndrome shows only incomplete penetrance, we would get discordance.

TABLE 1

Numbers of males scored for all five rDNA bands (A, B, C, D and G) and, of those, the number successfully testcrossed to females of the *aa* tester shock

Year	No. scored for rDNA	No. testcrossed
1983	185	166
1984	79	25
1986	257	220
1987–1988	515	425
Total	1036	836

RESULTS

The number of males scored and their year of collection is shown in Table 1. Some testcrosses did not take or did not result in sufficient numbers of progeny to score accurately for *aa* by morphological criterion. Moreover, some flies were frozen directly and were not testcrossed. Finally, some flies were successfully testcrossed, but either died before they could be frozen for DNA analysis, or the DNA extraction procedure did not work properly. These flies are not included in the data base. Table 1 also shows the sample sizes for males that were scored both by molecular and testcross criteria.

Figure 2 shows the estimated proportion of inserted 28S genes on the *X* chromosome for our sample of 1036 chromosomes. As is evident from that figure, considerable genetic variation for insert proportion exists in the natural population. The natural population contains *X* chromosomes that range from being virtually devoid of R1 inserts (the low end of the range of the estimator is 0.057) to having no detectable noninserted 28S genes on the *X* (the upper end of the range of the estimator is 1.000). The average *X*-chromosome has 55.7% of its rDNA repeats bearing the R1 insert, and the median is 56.1%. The variance in insert proportion is 0.0180. As mentioned in the introduction, studies with laboratory stocks indicate that having roughly a third or more of the *X*-linked 28S genes with the insert is necessary for the morpho-

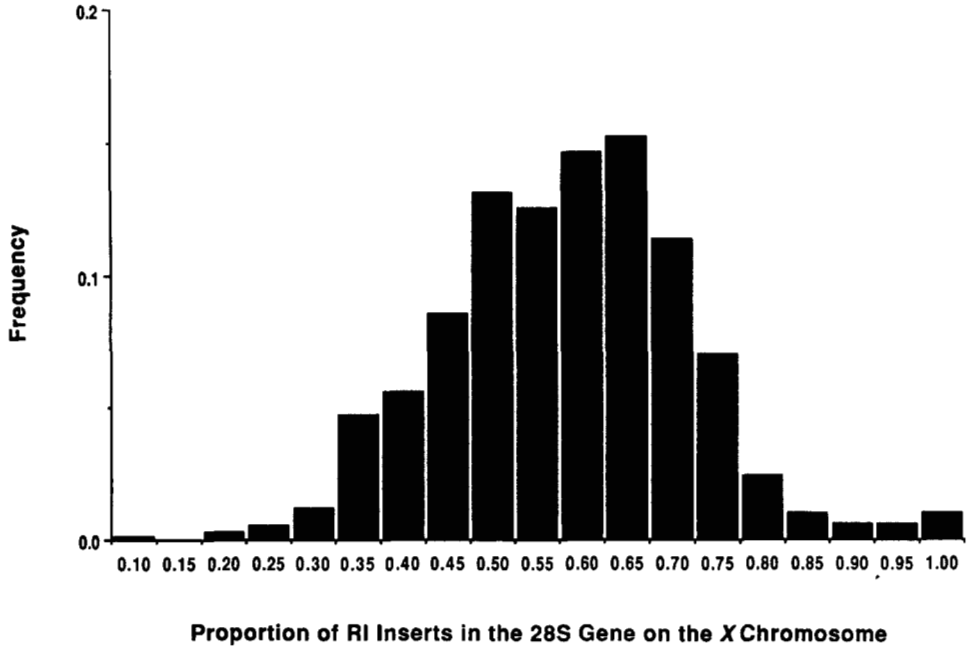

FIGURE 2.—The proportion of 28S rDNA repeats that bear the R1 insert in 1036 *X* chromosomes extracted from a natural population of *D. mercatorum* living near Kamuela, Hawaii.

logical expression of *aa* under laboratory conditions. Using this criterion, 94.9% of the *X* chromosomes in the natural population are above this one-third threshold value.

Of the 836 males for which testcross information also exists, 304 had *X* chromosomes that were scored as being *aaX* by the criteria given earlier. This implies that the underreplication locus is also polymorphic in the natural population with two allelic classes (*aa* an +) being present at intermediate frequencies. To see if there is any association between the testcrossed-inferred allelic states at the *ur* locus with the proportion of R1 inserted 28S genes, we plotted the distribution of insert proportions separately for the 304 *X* chromosomes scored as being *aa* at the *ur* locus and for the 532 wild-type *X* chromosomes. The results are shown in Figure 3. Figure 3 shows that the distributions of insert proportions is skewed in opposite directions for + *vs. aa X* chromosomes. The 532 wild-type *X* chromosomes had an average of 54.9% of their 28S genes bearing the R1 insert, with a variance of 0.0168 and a skew of −0.0138. In contrast, the 304 *aa X* chromosomes had a mean of 57.2%, a variance of 0.0174, and a skew of 0.223. To see if the difference in skew is significant, we applied the median test to these data, a nonparametric test for examining differences in skew. Under this test, we first identify the median of the entire distribution for all *X* chromosomes, which is 56.1% as stated above. We then count up the number of + and *aa X* chromosomes that are above and below the overall median. The results are shown in Table 2. The null hypothesis of no difference in skew is now tested by a standard chi-square contingency statistic on this 2 by 2 table. The resulting test statistic is 4.65 with 1 degree of freedom, which is significant at the 5% level. Hence, we reject the null hypothesis. This significant difference in the distribution of inserted 28S repeats means that 97.4% of all *X* chromosomes that are *aa* at the *ur* locus are above the threshold value of 1/3 for expression of *aa* in the laboratory.

We also examined the correlations between insert proportions and penetrance and expression of *abnormal abdomen* in the female testcross progeny of the 304 males bearing *aa X* chromosomes. The results are given in Table 3. All the correlations are small and positive, but none are significant at the 5% level.

A total of 1036 males were scored for the presence or absence of the D band, and of these, 834 yielded sufficient testcross progeny to infer expression of *aa* in their male offspring. Some 16% of all *Y* chromosomes in the sample of 1036 males lacked the D band. To see if the presence or absence of the D band was associated with the expression of *aa* in the male testcross progeny, we divided the sample of 834 males into two categories: + *Y* males and *aa Y* males based on the testcross results. 14.6% of the + *Y* chromosomes lacked D bands, whereas 26.4% of the *aa Y* chromosomes lacked D bands. The numbers of *Y* chromosomes with and without D bands in these two testcross categories are given in Table 4. The contingency chi-square statistic from this table is 12.15 with 1 d.f., which is significant at the 0.1% level.

To test for the concordance of *aa* penetrance in

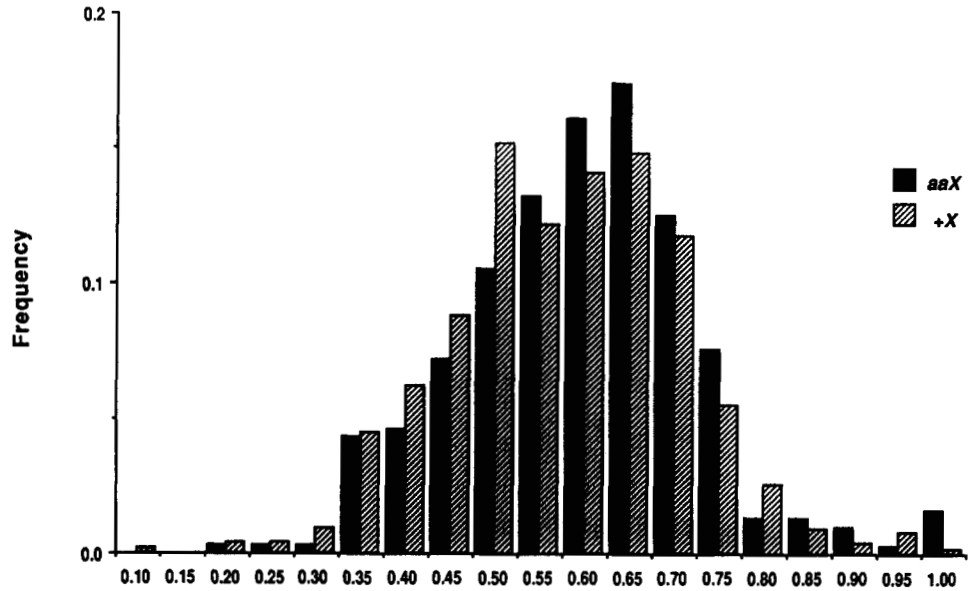

FIGURE 3.—The proportions of 28S rDNA repeats that bear the R1 insert in 532 *X* chromosomes characterized as being + at the *ur* locus by testcross results and in 304 *X* chromosomes characterized as being *aa* at the *ur* locus by testcross results. All *X* chromosomes are extracted from a natural population of *D. mercatorum* living near Kamuela, Hawaii.

TABLE 2

Numbers of + and *aa* X chromosomes with their percentage of inserted 28S genes above and below the overall median value of 56.2%

Chromosomes	Below median	Above median
+ *X*	281	251
aa X	137	167

An *X* chromosome is inferred to be either + or *aa* at the *ur* locus on the basis of the testcross data.

TABLE 3

Pearson correlation coefficients of the proportion of inserted 28S genes with the penetrance, expressivity, and penetrance times expressivity of *abnormal abdomen* in the sample of 304 X chromosomes that were inferred to be *aa* at the *ur* locus

Variable	Correlation coefficient
Penetrance	0.072
Expressivity	0.026
Penetrance × expressivity	0.080

The significance of the correlation coefficients are obtained by treating $p/(1 - p^2)$ as coming from a t distribution with $n - 2$ degrees of freedom, where $n = 304$ and p is the appropriate correlation. None of the correlations were significantly different from 0.

male *vs.* female testcross progeny, we scored *aa* expression separately for the sons and daughters of 268 wild-caught females crossed to the *aa* stock. If *aa* is completely penetrant in both sexes, we expect complete concordance for sons and daughters, but if this syndrome shows only incomplete penetrance, we expect discordance. Table 5 shows the results. As expected, there is a highly significant association between sons and daughters for *aa* expression ($\chi^2 = 63.25$ with 1 d.f., $P < 10^{-5}$), but there is not complete concordance. Both types of discordance occur; that is, in some cases the daughters express *aa* but the sons do not, while in other cases the sons express *aa* but the daughters do not. Hence, there is incomplete penetrance in both females and males. However, the degree of incomplete penetrance is not equal in the

TABLE 4

The number of *Y* chromosomes with and without D bands from males giving rise to male testcross progeny with (*aa Y*) and without (+ *Y*) expression of *abnormal abdomen*

Chromosomes	With D band	Without D band
aa Y	109	39
+ *Y*	586	100

TABLE 5

The expression of *abnormal abdomen* phenotypes in the daughters and sons of wild-caught females mated to *aa* stock males

	Daughters	
Sons	+	*aa*
+	153	55
aa	10	50

Each sex was scored separately for *aa* using the criterion of two or more affected individuals out of a total of 50 siblings of the same sex.

two sexes. Given the presence of *aa X* chromosomes in the wild-caught mother as inferred by *aa* expression in the daughters and/or sons, the probability of no *aa* expression in the daughters is $p_1 = 10/115 = 0.09$ (from Table 5). In contrast, the probability of no *aa* expression in the sons given an *aaX* and a known *aaY* is $p_2 = 55/115 = 0.48$. We tested the significance of this difference through the statistic:

$$(p_1 - p_2)/\{[p_1(1 - p_1) + p_2(1 - p_2) + 2p_1p_2]/115\}^{1/2} \quad (2)$$

which has an asymptotic normal distribution with mean 0 and variance 1 under the null hypothesis that $p_1 = p_2$. The value of this statistic is -6.54, for which $P < 10^{-5}$ under a two-tailed test of the null hypothesis. Hence, *aa* expression is far more penetrant in females than it is in males.

DISCUSSION

Previous studies on laboratory strains revealed three major genetic and molecular components to *abnormal abdomen* expression in *D. mercatorum*: R1 element insertions in the *X*-linked rDNA, the absence of replicative selection against inserted repeats in polytene tissue, and the absence of rDNA on the *Y* chromosomes (TEMPLETON, CREASE and SHAH 1985; DESALLE, SLIGHTOM and ZIMMER 1986; DESALLE and TEMPLETON 1986). The studies reported in this paper clearly quantify the level of variation for all three of these major components of *abnormal abdomen* in a natural population living near Kamuela, Hawaii.

The extent of variation in insert proportions is particularly remarkable. The natural variation spans virtually the entire range of theoretically possible values, thereby showing that rDNA multigene families coexist in the same population that differ almost as much as maximally possible in their proportion of inserted repeats. This result contrasts with the low variation in proportion of inserted repeats found in a survey of 96 *X* chromosomes in *D. melanogaster* (LYCKEGAARD and CLARK 1991). Variation within multigene families is influenced by many mechanisms that operate at the molecular level, such as unequal recombination, transposition and gene conversion (sometimes collectively called "molecular drive") (DOVER 1982). DOVER (1982) has also argued that these molecular-level factors can homogenize the repeats within a multigene family so rapidly that little intraspecific variation is expected and that therefore little opportunity for natural selection exists on the variants within a multigene family. However, theoretical models have shown that a pattern of low amounts of intraspecific variation is expected only under certain parameter values, whereas other parameter values can result in large amounts of intraspecific variation within the multigene family (OHTA and DOVER 1983; SLATKIN 1986; WILLIAMS 1990). Obviously, rDNA in *D. mercatorum* falls into this latter category of parameter values, whereas *D. melanogaster* appears to fall within the former category. Consequently, contrary to predictions that "molecular drive" can minimize the role of natural selection in multigene families (DOVER 1982), our results indicate that there is considerable opportunity for natural selection to operate upon the rDNA of *D. mercatorum*. Initial analyses strongly indicate that this opportunity is actually being realized (TEMPLETON *et al.* 1989); later papers in this series will address this issue more directly.

The studies reported here have mixed implications for the potential adaptive role of variation in R1 insert proportion. Given the one-third threshold value for *aa* expression, the distributions shown in Figures 2 or 3 imply that although insert variation is significant, it is not a major contributor to the presence or absence of the *abnormal abdomen* expression in nature simply because the vast majority of all *X* chromosomes are well above this threshold value. However, this threshold value is based on morphological expression only under laboratory conditions. It is not at all clear if this threshold value is valid for all the other phenotypes associated with this syndrome or with expression under natural conditions. The low correlations shown in Table 3 also imply that insert proportion variation is not a major contributor to penetrance or expressivity given the presence of *abnormal abdomen*. Given these results, it appears that the *ur* locus, which controls differential polytenization, behaves as the major determinant of the *aa* syndrome, and the frequency of the R1 insert behaves as a modifier locus with respect to *ur*.

Studies on *D. hydei* suggest that a stronger correlation may exist between number (as opposed to proportion) of noninserted 28S genes with penetrance and expressivity. The *bobbed* syndrome in *D. hydei* is also due to insertions in the *X* linked 28S genes and bears many similarities to *abnormal abdomen* (TEMPLETON *et al.* 1989). The data given in FRANZ and KUNZ (1981) shows that when the proportion of inserted repeats exceeds 40%, *bobbed* phenotypes arise in *D. hydei* and that the severity of the *bobbed* phenotype is correlated with the proportion of inserted genes. FRANZ and KUNZ (1981) also estimated the numbers of noninserted genes as well as their proportions, and the correlations are much stronger with the numbers as opposed to the proportions. A similar situation may be true for *D. mercatorum*. However, estimates for the total number of rRNA genes, which are needed to convert the proportions into the number of noninserted repeats, are technically difficult and associated with large errors (BIRNSTIEL, CHIPCHASE, and SPEIRS, 1971; LYCKEGAARD and CLARK 1989). Because of the difficulty in estimating total number of repeats, pre-

vious surveys of this nature have involved only 40 (FRANZ and KUNZ 1981), 25 (LYCKEGAARD and CLARK 1989), and more recently, 240 (LYCKEGAARD and CLARK 1991) Drosophila chromosomes. For a natural survey on the scale of this study (over 1000 each of wild-extracted *X* and *Y* chromosomes analyzed), estimating the total number of repeats would be extremely laborious with present techniques. Because earlier studies of laboratory stocks of *D. mercatorum* covering the full range of severity in the *aa* syndrome showed no differences in total rDNA content (DESALLE, SLIGHTOM and ZIMMER 1986), we decided to sacrifice the finer scale resolution that may have been offered by using the number of noninserted repeats for the added statistical power offered by large sample sizes that were made possible by employing the proportion measure. Our increased sample size has paid off by revealing an association between the *ur* locus and the proportion of R1 inserts (see Figure 3), an association which would not have been revealed in a more detailed but smaller scale survey.

The association between the *ur* locus and the proportion of R1 inserts on the *X* chromosome supports the hypothesis that the variation in insert proportion is of some adaptive or phenotypic significance. The two primary genetic components that are necessary for *abnormal abdomen* expression in females tend to cosegregate in the natural population. The frequency of recombination between the *ur* locus and the 18S/28S rRNA gene complex has been estimated to be between 0.004 and 0.007 (TEMPLETON, CREASE and SHAH 1985). This recombination frequency is sufficiently large that no linkage disequilibrium would be expected in a large population under neutrality, but this recombination frequency is sufficiently small that linkage disequilibrium could easily build up if selection favored such an association. Hence, the existence of the association itself is indicative of some selective role of both the *ur* locus and the insert proportions. Subsequent papers in this series will address the issue of adaptive significance more directly. For the present, we emphasize that the association of the two necessary components for female *aa* expression coupled with the low recombination frequency imply that the *ur* locus and the 18S/28S rRNA region are behaving as a supergene for *abnormal abdomen* expression in females.

The association between the *ur* locus and the proportion of R1 inserts has implications for the nature of the R1 elements themselves, as well. An association between the *ur* locus and the 18S/28S gene cluster is not expected if the R1 elements are actively transposing, as has been suggested (XIONG and EICKBUSH 1988a,b; XIONG *et al.* 1988; JAKUBCZAK, BURKE and EICKBUSH 1991). If these R1 elements are transposing, then the rate of transposition must be on the same order of magnitude as the recombination frequency between the two loci; and, in this context, the fact that they may be transposing in the traditional sense may have little effect on the adaptive significance of the R1 elements.

A more interesting approach to the evolutionary history of the R1 elements may be to look at their role in gene regulation; especially, how they may function in the polytenization process itself. It is well recognized in a number of different insect systems that ribosomal DNA repeats can be selectively replicated in polytene tissues, the common scenario being that noninserted repeats are replicated, while the insert-containing repeats are not and often found to be associated with adjacent heterochromatin rather than the nucleolus organizer (BECKINGHAM 1982; BECKINGHAM and RUBACHA 1984; KORGE 1987; SPRADLING 1987). Because sequence studies indicate that the only difference between inserted and noninserted repeats is the presence of the R1 element (LONG *et al.* 1981; MANDAL and DAWID 1981; SMITH and BECKINGHAM 1984), the control for differential replication must interact with the element itself. In addition, the same sequence data indicates no structural reason for the inserted repeats not to be transcribed. In fact, in a cell-free transcription system, KOHORN and RAE (1982) found no qualitative differences in transcription between inserted and noninserted rDNA genes. Hence, differential replication and transcriptional control are both mediated by the presence of the R1 element alone. The commonly accepted explanation for these observations is that replication and transcription levels are directly affected by the physical state of the chromatin. More highly condensed chromatin is not accessible by proteins involved in these two molecular processes. The decreased replication and transcription of inserted rDNA repeats results when these repeats become a part of the more highly condensed heterochromatin in polytene tissues (BECKINGHAM and RUBACHA 1984; EISSENBERG 1989; KARPEN and SPRADLING 1990). What has not been decided is the mechanism of heterochromatin condensation itself in polytene tissues, and especially how specific subsets of the rDNA array are selected for condensation and others are not. KARPEN and SPRADLING (1990) have recently proposed a new model for differential replication which involves transposable element-mediated somatic elimination of DNA to create underreplicated chromatin. If their model proves to be correct, then the R1 elements may be actively excising during the polytenization process leading to the observed underreplication and lack of transcription for insert-bearing rDNA repeats. Therefore, the R1 elements may be actively transposing afterall, not in the traditional sense, but rather in a more tissue-specific, restricted manner related to the polyteniza-

tion process. In the *abnormal abdomen* system of *D. mercatorum*, the *ur* locus can then be seen as a more general control element for the transposition of R1 and R1-associated elements in polytene tissues.

We also investigated *Y* chromosomes to see if there is any genetic variation for *aa* expression in males in the natural population. We found that there is extensive polymorphism for the presence *vs.* absence of the *Y*-specific D band, with 17% of all *Y* chromosomes in nature lacking the D band. This polymorphism for the D band explains much, but not all, of the phenotypic expression of *abnormal abdomen* in the testcross sons (Table 4). It can be seen from Table 4 that *Y* chromosomes lacking the D band do not always express *aa* in males. One possible explanation for these anomalous *Y* chromosomes is that they contain *Y*-linked rDNA but with spacers that are the same size as those found on the *X*. In this case, the absence of the D band would not correspond to the absence of *Y*-linked rDNA. We do not believe this is the case, because previous population surveys of restriction site variation of *X* and *Y* chromosome rDNA in *D. mercatorum* and *D. melanogaster* have revealed no instances of *X* and *Y* rDNA repeats sharing the same nontranscribed spacer lengths (INDIK and TARTOFF 1980; WILLIAMS, DESALLE and STROBECK 1985; WILLIAMS *et al.* 1987; GILLINGS *et al.* 1987). Hence, distinct nontranscribed spacer lengths appears to be the trademark for Drosophila *X* and *Y* chromosome rDNA. The rate of interchromosomal exchange which potentially could homogenize repeat types between the *X* and *Y* chromosome rDNA is low (GILLINGS *et al.* 1987; LYCKEGAARD and CLARK 1991), and when it does occur, interchromosomal recombination results in the transfer of *Y*-specific sequences to the *X* chromosome rather than *X* sequences to the *Y* chromosome because of the unstable nature of the resulting recombination intermediates (GILLINGS *et al.* 1987; ENGLAND, STOKES and FRANKHAM 1988; FRANKHAM 1990). Therefore, we believe that the absence of the D band is a good indicator for a lack of rDNA on the *Y* chromosome. A more likely explanation for observing *Y* chromosomes that lack the D band, yet do not express *aa* in males is the low penetrance of *aa* in male testcross progeny (Table 5). Because of this we would expect to have many cases of the *aaY* chromosomes being scored as *+Y* chromosomes from the testcross data. This low penetrance could account for most of the flies lacking a D band that were scored as *+Y* on the basis of the testcross (the lower right-hand category in Table 4). In addition, there could be autosomal modifiers present that suppress the morphological expression of *aa* in males. As reported in TEMPLETON, CREASE and SHAH (1985), autosomal suppressors of *aa* expression in females exist. That paper dealt only with female expression, but unpublished data clearly indicate that male expression can also be modified. Recall that the male parent of the testcross progeny is either a wild-caught male or the son of a wild-caught, naturally inseminated female. Hence, we have no control over the autosomal genetic background brought in by the male parent. Accordingly, if the natural population is polymorphic for autosomal modifiers of *aa* expression, it would reduce the association between *Y* chromosomal variants and male *aa* expression (Table 4) as well as the association between *X* chromosomal variants and female *aa* expression (Table 3).

Table 4 also reveals the existence of *Y* chromosomes with D bands that nevertheless allow *aa* expression in males. These anomalies might also be due to autosomal modifiers, but they may be due to additional *Y* chromosomal polymorphism. These *Y* chromosomes may indeed have rDNA (thereby explaining the presence of the D band), but the amount of *Y*-linked rDNA might be very low and insufficient to compensate for an *X* chromosome that bears both the *aa* allele at the *ur* and with 76% of the 28S genes bearing the R1 insert. In addition to quantitative variation in the amount of *Y*-linked rDNA, there may be variation in the *Y* chromosomes for selective underreplication or nucleolar dominance with respect to the *X* chromosome. The *Y* chromosome studied by DeSalle and TEMPLETON (1986) that had the D band and suppressed *aa* expression was greatly overreplicated in polytene tissues compared to *aa X* chromosomes. However, if some *Y* chromosomes are not favored over *aa Xs* by selective underreplication or do not show nucleolar dominance over the *X* chromosome in polytene tissues, then there would be no suppression of *aa* expression in males.

It is obvious from our survey that *Y* chromosomes are highly polymorphic in this natural population of *D. mercatorum*, and that phenotypic differences can be ascribed to this variation. CLARK (1987) has shown that it is difficult for constant fitness selection models to maintain *Y* polymorphisms. HENCE, CLARK (1987) concluded that "either *Y*-linked variation is neutral, that it involves frequency dependent fitnesses or geographical structuring, or that it interacts with another chromosome." All of these alternatives are certainly possibilities in this case. The natural population being studied lives in a highly heterogeneous environment with limited dispersal (JOHNSTON and TEMPLETON 1982) and shows significant geographical structuring for mitochondrial DNA variation (DESALLE *et al.* 1987) and for the *X*-linked components of *abnormal abdomen* (TEMPLETON *et al.* 1989). Moreover, with respect to the phenotype of male morphological expression, we know that the *Y* chromosomal polymorphism is strongly interacting with the *X* chromosome and to a lesser extent with the autosomes. Therefore, natural selection cannot be ruled out as a possible

evolutionary force shaping rDNA variation on the *Y* chromosome of *D. mercatorum*.

The survey work presented in this paper clearly demonstrates that all three major molecular components of *abnormal abdomen* (selective underreplication, R1 inserts, and *Y* chromosomes with and without D bands) are highly polymorphic in the natural population of *D. mercatorum* that lives near Kamuela, Hawaii. Moreover, these studies indicate that the two *X*-linked components are nonrandomly associated and effectively form a supergene complex that controls the expression of *aa* in females. Given the existence of this variation, we will turn our attention to the phenotypic and adaptive significance of this variation in subsequent papers in this series.

We wish to thank ROSE LEIGHTNER for technical help that was essential to this work. We also wish to thank two anonymous reviewers for their helpful suggestions for improving this manuscript. This research was supported by National Institutes of Health grant R01 AG02246, and by National Institutes of Health Genetics Training grant GM08036 and an American Association of University Women American Fellowship awarded to H.H.

LITERATURE CITED

BACK, E., E. VAN MEIR, F. MULLER, D. SCHALLER, H. NEUHAUS, P. AEBY and H. TOBLER, 1984 Intervening sequences in the ribosomal RNA genes of *Ascaris lumbricoides*: DNA sequence at the junctions and genomic organization. EMBO J. **3:** 2523–2529.

BECKINGHAM, K., 1982 Insect rDNA, pp. 206–269 in *The Cell Nucleus*, Vol. X, edited by H. BUSCH and L. ROTHBLUM. Academic Press, New York.

BECKINGHAM, K., and A. RUBACHA, 1984 Different chromatin states of the intron$^-$ and type I intron$^+$ rRNA genes of *Calliphora erythrocephala*. Chromosoma **90:** 311–316.

BECKINGHAM, K., and R. L. WHITE, 1980 The ribosomal DNA of *Calliphora erythrocephala*; an analysis of hybrid plasmids containing ribosomal DNA. J. Mol. Biol. **137:** 349–373.

BIRNSTIEL, M. L., M. CHIPCHASE and J. SPEIRS, 1971 The ribosomal cistrons. Prog. Nucleic Acid Res. Mol. Biol. **11:** 351–389.

BURKE, W. D., C. C. CALALANG and T. H. EICKBUSH, 1987 The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. **7:** 2221–2230.

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

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.

DESALLE, R., J. SLIGHTOM and E. Zimmer, 1986 The molecular through ecological genetics of *abnormal abdomen* in *Drosophila mercatorum*. II. Ribosomal DNA polymorphism is associated with the *abnormal abdomen* syndrome in *Drosophila mercatorum*. Genetics **112:** 861–875.

DESALLE, R., and A. R. TEMPLETON, 1986 The molecular through ecological genetics of *abnormal abdomen* in *Drosophila mercatorum*. III. Tissue-specific differential replication of ribosomal genes modulates the *abnormal abdomen* phenotype in *Drosophila mercatorum*. Genetics **112:** 877–886.

DESALLE, R., A. R. TEMPLETON, I. MORI, S. PLETSCHER and J. S. JOHNSTON, 1987 Temporal and spatial heterogeneity of mtDNA polymorphisms in natural populations of *Drosophila mercatorum*. Genetics **116:** 215–223.

DOVER, G.A., 1982 Molecular drive: a cohesive mode of species evolution. Nature **299:** 111–117.

EICKBUSH, T. H., and B. ROBINS, 1985 *Bombyx mori* 28S ribosomal genes contain insertion elements similar to the type I and II elements of *Drosophila melanogaster*. EMBO J. **4:** 2281–2285.

EISSENBERG, J. C., 1989 Position-effect variegation in *Drosophila*: towards a genetics of chromatin assembly. Bioessays **11:** 14–17.

ENGLAND, P. R., H. W. STOKES and R. FRANKHAM, 1988 Clustering of rDNA containing type 1 insertion sequence in the distal nucleolus organiser of *Drosophila melanogaster*: implications for the evolution of X and Y rDNA arrays. Genet. Res. **51:** 209–215.

FRANKHAM, R., 1990 Adding the heterochromatic Y^L arm to an X chromosome reduces reproductive fitness in *Drosophila melanogaster*: implications for the evolution of rDNA, heterochromatin, and reproductive isolation. Genome **33:** 340–347.

FRANZ, G., and W. KUNZ, 1981 Intervening sequences in ribosomal RNA genes and *bobbed* phenotype in *Drosophila hydei*. Nature **292:** 638–640.

FRANZ, G., W. KUNZ and C. GRIMM, 1983 Determination of the region of the rDNA involved in polytenization in salivary glands of *Drosophila hydei*. Mol. Gen. Genet. **191:** 74–80.

GILLINGS, M. R., R. FRANKHAM, J. SPEIRS and M. WHALLEY, 1987 *X-Y* exchange and the coevolution of the *X* and *Y* rDNA arrays in *Drosophila melanogaster*. Genetics **116:** 241–251.

GOODRICH-YOUNG, C., and H. M. Krider, 1989 Nucleolar dominance and replicative dominance in *Drosophila* interspecific hybrids. Genetics **123:** 349–358.

INDIK, Z. K., and K. D. TARTOF, 1980 Long spacers among ribosomal genes in *Drosophila melanogaster*. Nature **284:** 477–479.

JAKUBCZAK, J. L., W. D. BURKE and T. H. EICKBUSH, 1991 Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA **88:** 3295–3299.

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

JOHNSTON, J. S., and A. R. TEMPLETON, 1982 Dispersal and clines in Opuntia breeding *Drosophila mercatorum* and *D. hydei* at Kamuela, Hawaii, pp. 241–256 in *Ecological Genetics and Evolution*, edited by J. S. F. BARKER and W. T. STARMER. Academic Press, New York.

KARPEN, G. H., and A. C. SPRADLING, 1990 Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in *Drosophila*. Cell **63:** 97–107.

KOHORN, B. D., and P. M. M. RAE, 1982 Accurate transcription of truncated ribosomal DNA templates in a *Drosophila* cell-free system. Proc. Natl. Acad. Sci. USA **79:** 1501–1505.

KORGE, G., 1987 Polytene Chromosomes, pp. 27–58 in *Structure and Function of Eukaryotic Chromosomes*, edited by W. HENNIG. Springer-Verlag, New York.

LECANIDOU, R., T. H. EICKBUSH and F. C. KAFATOS, 1984 Ribosomal DNA genes of *Bombyx mori*: a minor fraction of the repeating units contain insertions. Nucleic Acids Res. **12:** 4703–4713.

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

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

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

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

LYCKEGAARD, E. M. S., and A. G. CLARK, 1991 Evolution of ribosomal RNA gene copy number on the sex chromosomes of *Drosophila melanogaster*. Mol. Biol. Evol. **8:** 458–474.

MAKNI, M., M. MARRAKCHI and N. PRUD'HOMME, 1989 The occurrence of long ribosomal transcripts homologous to type I insertions in *bobbed* mutants of *Drosophila melanogaster*. Genet. Res. **54:** 127–135.

MANDAL, R. K., and I. B. DAWID, 1981 The nucleotide sequence at the transcription termination site of ribosomal RNA in *Drosophila melanogaster*. Nucleic Acids Res. **9:** 1801–1810.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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.

PASKEWITZ, S. M., and F. H. COLLINS, 1989 Site-specific ribosomal DNA insertion elements in *Anopheles gambiae* and *A. arabiensis*: nucleotide sequence of gene-element boundaries. Nucleic Acids Res. **17:** 8125–8133.

RITOSSA, F., 1976 The *bobbed* locus, pp. 801–846 in *The Genetics and Biology of Drosophila*, Vol 1b, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.

ROIHA, H., and D. M. GLOVER, 1980 Characterisation of complete type II insertions in cloned fragments of ribosomal DNA from *Drosophila melanogaster*. J. Mol. Biol **140:** 341–355.

SLATKIN, M., 1986 Interchromosomal biased gene conversion, mutation and selection in a multigene family. Genetics **112:** 681–698.

SMITH, V. L., and K. BECKINGHAM, 1984 The intron boundaries and flanking rRNA coding sequences of *Calliphora erythrocephala* DNA. Nucleic Acids Res. **12:** 1707–1724.

SOUTHERN, E., 1979 Gel electrophoresis of restriction fragments. Methods Enzymol. **68:** 152–176.

SPRADLING, A., 1987 Gene amplification in Dipteran chromosomes, pp. 199–212 in *Structure and Function of Eukaryotic Chromosomes*, edited by W. HENNIG. Springer-Verlag, New York.

SPRADLING, A., and T. ORR-WEAVER, 1987 Regulation of DNA replication during *Drosophila* development. Annu. Rev. Genet. **21:** 373–403.

TEMPLETON, A. R., 1979 The parthenogenetic capacities and genetic structures of sympatric populations of *Drosophila mercatorum* and *Drosophila hydei*. Genetics **92:** 1283–1293.

TEMPLETON, A. R., 1982 The prophecies of parthenogenesis, pp. 75–101 in *Evolution and Genetics of Life Histories*, edited by H. DINGLE and J. P. HEGMANN. Springer-Verlag, New York.

TEMPLETON, A. R., 1983 Natural and experimental parthenogenesis, pp. 343–398 in *The Genetics and Biology of Drosophila*, Vol. 3C, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press, London.

TEMPLETON, A. R., T. J. CREASE and F. SHAH, 1985 The molecular through ecological genetics of *abnormal abdomen* in *Drosophila mercatorum*. I. Basic genetics. Genetics **111:** 805–818.

TEMPLETON, A. R., and L. R. LAWLOR, 1981 The fallacy of the averages in ecological optimization theory. Am. Nat. **117:** 390–393.

TEMPLETON, A. R., and M. A. RANKIN, 1978 Genetic revolutions and control of insect populations, pp. 83–112 in *The Screwworm Problem*, edited by R. H. RICHARDSON. University of Texas Press, Austin.

TEMPLETON, A. R., H. HOLLOCHER, S. LAWLER and J. S. JOHNSTON, 1989 Natural selection and ribosomal DNA in *Drosophila*. Genome **31:** 296–303.

TERRACOL, R., 1986 Transcription of rDNA insertions in *bobbed* mutants of *Drosophila melanogaster*. Genet. Res. **48:** 167–174.

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

WELLAUER, P. K., and I. B. DAWID, 1977 The structural organization of ribosomal DNA in *Drosophila melanogaster*. Cell **10:** 193–212.

WELLAUER, P. K., and I. B. DAWID, 1978 Ribosomal DNA in *Drosophila melanogaster*. II. Heteroduplex mapping of cloned and uncloned rDNA. J. Mol. Biol. **126:** 769–782.

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.

WHITE, R. L., and D. S. HOGNESS, 1977 R loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. Cell **10:** 177–192.

WILLIAMS, S. M., 1990 The opportunity for natural selection on multigene families. Genetics **124:** 439–441.

WILLIAMS, S. M., R. DESALLE and C. STROBECK, 1985 Homogenization of geographical variants at the nontranscribed spacer of rDNA in *Drosophila mercatorum*. Mol. Biol. Evol. **2:** 338–346.

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 the *X* and *Y* chromosomes. Genetics **116:** 225–232.

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

XIONG, Y., and T. H. EICKBUSH, 1988b Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell **55:** 235–246.

XIONG, Y., W. D. BURKE, J. L. JAKUBCZAK and T. H. EICKBUSH, 1988 Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes. Nucleic Acids Res. **16:** 10561–10573.

Communicating editor: D. CHARLESWORTH

APPENDIX

TEMPLETON *et al.* (1989) used the estimator G/(A − D + G). The quantity A − D should ideally be proportional to the number of noninserted *X*-linked repeats, and G should be proportional to the number of inserted *X*-linked repeats. However, this estimator assumes that all three bands yield comparable signals upon autoradiography. In that regard, it is important to note that our probe hybridizes well only to the coding regions of the rDNA repeats. SOUTHERN (1979) pointed out that the efficiency of hybridization falls off sharply at and below sizes of about 1 kb. The B, C and D bands all are associated with hybridization with DNA of about 1 kb in size, whereas the A and G peaks are associated with hybridization with much larger DNA fragments. Hence, one would expect that the intensity of the signals associated with these two sets of bands would not be comparable. That this is indeed the case is shown clearly by a contrast of the peak heights of the G and C bands. The DNA fragments associated with these bands should be equally abundant (Figure 1). Hence, the statistic G − C should have a zero expectation. In a sample of 1044 *X* chromosomes, the average G − C difference was 5.3 mm. This mean value was significantly different from zero using a paired *t*-test (t = 58.3 with 1043 d.f., $P < 10^{-6}$). Moreover, only one observation out of the 1044 was less than zero (−0.9). It is therefore obvious that the hybridization signal associated with the G band is much more intense than that associated with the C band, which results

in a serious underestimate of the true amount of the C band in the filter relative to the G band. Since the D band is visualized through hybridization with a DNA segment of comparable size to the C band, the G − C difference implies that subtracting off the D band peak height from that of the A band will not adequately correct for the rDNA found on the *Y* chromosome. Consequently, we need to devise a correction factor to adjust for the hybridization differences (FRANZ, KUNZ and GRIMM 1983) or use an estimator that is based exclusively upon proportions of peak heights that are expected to yield comparable signal strengths.

The previous estimator (TEMPLETON *et al.* 1989) depended primarily upon the A and G bands, which are the most accurately measured peak heights. To use the A and G bands to estimate the proportion of *X*-lined inserted repeats, we need to correct for the fact that some of the A band intensity is due to rDNA genes on the *Y* chromosome. From Figure 1, note that (B − C)/(B − C + D) should estimate the proportion of A bands that are associated with *X*-linked repeats. This proportion is based upon bands that have nearly equally sized DNA segments that hybridize with the probe. This estimator will only make biological sense if B > C, an inequality that is expected from Figure 1. Out of 1041 flies scored for both bands, this inequality was satisfied in every case, and in only three cases did these two bands give equal peak heights. Hence, the relative peak heights of these two bands obey the biological restrictions implied by Figure 1, whereas bands G and C do not. We will therefore use this ratio to correct for *Y*-linked rDNA instead of subtracting the D band. Thus, an alternative estimator for the proportion of inserted *X*-linked repeats is:

$$G/\{((B - C)/(B - C + D))(A + G)\}. \quad (1)$$

Estimator (1) uses more of the data, but in such a way that it should not be affected by signal intensity differences between bands B, C and D *vs.* bands A and G.

The relationship of the bands in estimator (1) appears complicated. It is usually desirable to keep such estimators as simple as possible. The simplest estimator for the proportion of inserted *X*-linked repeats for our data would actually be C/B, but this estimate depends exclusively on the reliability of the measurement of the C band which suffers a double problem of having the weakest hybridization signal and of being the most diffuse of the bands because of the greater distance it must travel through the agarose gel. To ensure that we have not overlooked the possibility of being able to use this simpler estimator, we performed two sets of experiments to evaluate the relative merits of estimator (1) *vs.* the estimator C/B. The first set of experiments involved two highly inbred stocks, the *aa* tester stock and S-1-Br16. These two stocks were chosen for two reasons. First, they are highly inbred, so measurements on different individuals should reflect only measurement error and not genetic differences. Second, the *aa* stock has a very high proportion of its *X*-linked genes bearing the R1 insert, whereas the S-1-Br16 stock has a very low proportion. These two stocks span the range of almost all the naturally occurring variation. Hence, they are ideal for testing the properties of the alternative estimators over the relevant range of variation. For a sample of 33 individuals, estimator (1) yields an average R1 insert proportion of 0.19 for *X* chromosomes extracted from the S-1-Br16 stock, with a standard deviation of 0.14. In contrast, the average value of the C/B estimator is 0.74 with a standard deviation of 0.51. For 10 individuals from the *aa* stock, the mean value of estimator (1) is 0.76 with a standard deviation of 0.07, whereas the mean value of the C/B estimator is 0.94 with a standard deviation of 0.35. Hence, the measurement error associated with the C/B estimator is substantially higher than that of estimator (1) at both ends of the range of variability. Moreover, the estimators yield very different mean values. In principle, both estimators are biased since they incorporate the measured peak heights into nonlinear equations (TEMPLETON and LAWLOR 1981). However, since we do not know the true values, it is impossible to infer statistically which of the estimators is the more biased. We emphasize that the values associated with estimator (1) are much more biologically reasonable. The S-1-Br16 stock does not allow expression of the *aa* syndrome, and the G band is very weak on the autoradiographs associated with this stock. Thus, the value of 0.19 given by estimator (1) is more reasonable than the very high value of 0.74 given by C/B. Similarly, the C/B estimator for the *aa* stock is 0.94, which would indicate that very little functional rDNA would exist in this stock. Thus, both in terms of biological reasonableness and in terms of measurement error, estimator (1) is superior to the C/B estimator.

The second experiment involved Southern analyses on 69 wild-caught males. Three autoradiographs were developed from each filter; one with an exposure time of 7 hr, one of 10.5 hr, and one of 14 hr. These different exposure times result in different overall band intensities, and hence provide a test for the linearity of these estimators as a function of variation in overall band intensity. A two-way analysis of variance (ANOVA) was performed for each of the two estimators, with the treatments being an individual effect and an exposure time effect. For estimator (1), the individual effect was highly significant ($F = 13.0$ with 68 and 136 d.f., $P < 0.0001$), and the exposure time was significant at the 5% level ($F = 3.21$ with 2 and 136 d.f., $P < 0.0434$). For the C/B estimator, individual differences were significant at the 1% level ($F = 1.70$ with 68 and 136 d.f., $P < 0.0046$), and exposure time was highly significant ($F = 20.62$ with 2 and 136 d.f., $P < 0.0001$). Hence, for the C/B estimator, exposure time is a greater source of variation than individual differences, whereas for estimator (1), individual differences have a much larger effect than exposure time. Moreover, multiple comparisons tests (Tukey's studentized range test or Bonferroni tests) revealed significant differences between exposure time treatments for the C/B estimator, with the estimator increasing in value with decreasing exposure time such that the difference between the lowest and highest exposure times is 0.083. For estimator (1), there were no significant differences between the mean values at any of the exposure times, although the significance of exposure time in the ANOVA is due to an overall trend toward higher values with increasing exposure time, with the difference between the highest and lowest exposure times being 0.016. Thus, overall intensity has only a minor influence on estimator (1), but it is a major source of error for the C/B estimator. In light of these two experiments, the C/B estimator, despite its simplicity, is clearly inferior to estimator (1). Hence, we use equation (1) to estimate the proportion of *X*-linked rDNA repeats that bear the R1 insert.