Genet. Res., Camb. (1990), **55**, *pp.* 165–170 *With* 1 *text-figure Printed in Great Britain*

Segregation ratios within *Segregation Distorter* lines of *Drosophila melanogaster* conform to a beta-binomial distribution

CECIL B. SHARP* AND ARTHUR J. HILLIKER

Department of Molecular Biology and Genetics, University of Guelph, Guelph, Ontario N1G 2W1 Canada

(*Received 18 September 1989 and in revised form 20 February 1990*)

Summary

Segregation Distorter (*SD*) chromosomes are preferentially recovered from SD/SD^+ males due to the dysfunction of sperm bearing the SD^+ chromosome. The proportion of offspring bearing the *SD* chromosome is given the symbol *k*. The nature of the frequency distribution of *k* was examined by comparing observed *k* distributions produced by six different *SD* chromosomes, each with a different mean, with *k* distributions predicted by two different statistical models. The first model was one where the *k* of all males with a given *SD* chromosome were considered to be equal prior to the determination of those gametes which produce viable zygotes. In this model the only source of variation of *k* would be binomial sampling. The results rigorously demonstrated for the first time that the observed *k* distributions did not fit the prediction that the only source of variation was binomial sampling. The next model tested was that the prior distribution of segregation ratios conformed to a beta distribution, such that the distribution of *k* would be a beta-binomial distribution. The predicted distributions of this model did not differ significantly from the observed distributions of *k* in five of the six cases examined. The sixth case probably failed to fit a beta-binomial distribution due to a major segregating modifier. The demonstration that the prior distribution of segregation ratios of *SD* lines can generally be approximated with a beta distribution is crucial for the biometrical analysis of segregation distortion.

1. Introduction

The segregation distorter phenotype is caused by certain types of second chromosomes that can be recovered from natural populations (Sandler *et al.* 1959; Hiraizumi & Crow, 1960). These *Segregation Distorter* (*SD*) chromosomes cause a distortion of the observed segregation ratio when an *SD* chromosome is heterozygous with a normal chromosome 2. An SD/SD^+ male (but not an SD/SD^+ female) produces far more offspring bearing the *SD* chromosome than offspring with the SD^+ chromosome. The segregation ratio is usually defined in terms of the proportion of offspring inheriting the *SD* chromosome and is given the symbol *k*. SD/SD^+ males generally have mean *k* values far greater than 0·5.

Segregation Distorter chromosomes cause distortion of the segregation ratio by a genetically determined defect of spermiogenesis which results in the dysfunction of those spermatids bearing the SD^+ chromosome (Nicoletti, 1968; Tokuyasu *et al.* 1977). A strong *SD* chromosome is effectively transmitted as if it were homozygous, even though it is only heterozygous. (Since the effect is brought about prezygotically, there is very little reduction in fertility, as male *D. melanogaster* can produce a large excess of sperm.) There are several genetic elements on *SD* chromosomes that contribute to the full expression of the segregation distorter phenotype. The major loci are located in and near the centric heterochromatin of chromosome 2 (Hartl, 1974; Ganetzky, 1977; Martin & Hiraizumi, 1979; Brittnacher & Ganetzky, 1983, 1984; Sharp *et al.* 1985). Two of the elements act in *trans* to promote segregation distortion. Ganetzky (1977) has referred to them as *Sd* and *Enhancer of SD*, *E*(*SD*). The *Sd* element, when deleted from an *SD* chromosome, causes a very large reduction in *k*, while deletion of *E*(*SD*) causes a more moderate decrease in *k*. Sharp *et al.* (1985) have referred to these same genes as Sd_1 and Sd_2, since they observed that *SD* chromosomes that had Sd_1 removed by recombination retained significant albeit reduced distorting capacity, due to the presence of Sd_2 [*E*(*SD*)]. The other major locus that is important in segregation distortion is *Responder* (*Rsp*). This is a *cis*-acting genetic element which has alleles of varying sensitivity to the effects of the driving elements (Sd_1 and Sd_2) (Hiraizumi *et al.* 1980; Temin & Marthas, 1984; Lyttle *et al.* 1986).

* Present address: Department of Genetics, Stanford University, Stanford, CA 94305-5120, USA.

Sperm bearing chromosomes with sensitive *Rsp* alleles (*Rsp*^{s}) are more likely to be rendered dysfunctional than are sperm bearing insensitive alleles (*Rsp*^{i}). Some sensitive *Rsp* alleles have recently been cloned and it was shown that the degree of sensitivity was correlated with the number of copies of a tandemly repeated sequence (Wu *et al.* 1988). Another genetic element that is found on most naturally occurring *SD* chromosomes is *Modifier of SD* [*M*(*SD*)]. This is a dominant positive modifier of *SD* that is located in the proximal right arm of chromosome 2, near the *cinnabar* gene (Hiraizumi, Martin & Eckstrand, 1980).

Although segregation distortion is usually measured by the segregation ratio, *k*, it is apparent that the variance of *k* is highly dependent upon the mean of *k*, with the variance of *k* decreasing dramatically as the mean of *k* approaches 1·0 (Sandler & Hiraizumi, 1960). Miklos & Smith-White (1971) suggested a transformation of *k*, which we refer to as the *k*-probit transformation, which was intended to provide a proper interval measure for segregation distortion, thereby eliminating distortions of scale associated with the use of *k*. However, Sharp & Hilliker (1986) showed that observed frequency distributions of *k* values for a large data set did not always agree with the distributions predicted by the *k*-probit transformation. Furthermore, Sharp & Hilliker (1987) defined and critically assessed the assumptions of the *k*-probit transformation and found the assumptions to be numerous, and in several instances, unrealistic.

In the present report we will consider further the nature of variation of the segregation ratio among *SD* males. The first point that we will address is whether *k* is distributed as a binomial variable. Sandler & Hiraizumi (1961) were the first to suggest that the distribution of *k* does not conform to a binomial distribution, but they did not provide rigourous quantitative support for their assertion. Miklos & Smith-White (1971) and Sharp & Hilliker (1986) also assumed that there was extra-binomial variation, but neither of these reports demonstrated the statistical significance of the extra-binomial component. We will demonstrate here that there is indeed generally a significant amount of variation of the segregation ratio above that predicted by binomial sampling alone. The second point that this report will consider is what type of frequency distribution the extra-binomial component of variation conforms to. It will be shown that segregation ratios of *SD*/*SD*$^+$ males can generally be approximated by assuming a compound distribution of a binomial and a beta distribution (a beta-binomial distribution).

2. Materials and methods

A number of *SD* chromosomes were used that had different mean *k* values, in order to compare observed and predicted distributions over a range of conditions. A standard *SD* chromosome and five derivatives of that chromosome were employed. The derivation of the modified *SD* chromosomes is described in Sharp *et al.* (1985). The standard *SD* chromosome was *SD-5*. Four of the derivatives were recombinants between *SD-5* and a chromosome marked with *b pr lt pk cn*. *R*(*SD-5*)*pk cn* was a recombinant that has the *SD* genotype *Sd E*(*SD*) *Rsp*^{i} *M*(*SD*)$^+$, i.e. it has lost *M*(*SD*) by recombination and consequently has a reduced *k* value. *RR*(*SD-5*)*lt* was derived as a two-stage double recombinant that placed *lt* on the *SD* chromosome and removed *E*(*SD*), making its *SD* genotype *Sd E*(*SD*)$^+$ *Rsp*^{i} *M*(*SD*). *R*(*SD-5*)*b pr-5* is the product of a single exchange that has removed the *Sd* site, resulting in an *SD* genotype of *Sd*$^+$ *E*(*SD*) *Rsp*^{i} *M*(*SD*). It has a greatly reduced *k* value, but it has a slight amount of distortion, due to the distorting ability of *E*(*SD*). *RR*(*SD-5*)*pr lt* was another two-stage double recombinant. This chromosome had the *SD* genotype *Sd E*(*SD*)$^+$ *Rsp*^{ss} *M*(*SD*). *Df*(*2L*)*SD-5-8* had a deletion of the 2L heterochromatic gene *lt* induced in it by gamma rays. Although it was not deleted for *E*(*SD*) it likely had a reduced mean *k* value because of a position effect of the heterochromatic breakpoint upon the ability of the *E*(*SD*) locus to function.

The *SD* stocks were expanded from *SD*/*In*(*2LR*) *SM1* stock vials into bottles. Males from each *SD* line were collected and mated to *cn bw* virgin females in bottles. From this mating 1- or 2-day-old *SD*/*cn bw* males were individually mated in shell vials with two virgin females homozygous for the markers *cn bw*; *Ki p*^{p} *bx sr e*^{s}. Females with marked third chromosomes were used in order to ensure that the progeny of non-virgin females would not be scored as *SD*$^+$, which might have skewed the *k* distributions, given the large number of females collected. The parents were left in the shell vials for 5 days and then discarded. After all of the adults had eclosed the progeny were scored as either red eyed (*SD*/*cn bw*) or white eyed (*cn bw*/*cn bw*).

For each shell vial a *k* value was calculated as the number of *SD*-bearing progeny divided by the total number of progeny. The mean and variance of the *k* values from a given experiment were then computed without weighting. Although weighted descriptive statistics would have smaller standard errors, the theoretical contribution to precision would be negligible, because of the large sample sizes of these experiments, and consequently weighting would introduce unnecessary assumptions. In addition, the *k* values were grouped into classes with an interval width of 0·01, in order to determine the observed frequency distribution of *k*. The numbers of progeny produced by each male were also sorted into classes with an interval width of 1. Fig. 1 shows the frequency distribution of *k* for each chromosome, as well as the number of families scored and the average number of offspring per family.

We first wanted to compare the observed distributions with a binomial distribution. In order to do

https://doi.org/10.1017/S0016672300025489 Published online by Cambridge University Press

this there is a complication that must be considered: all males do not produce the same number of progeny. The general formula describing a binomial frequency distribution is

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},$$

where, in the context considered here, n is the number of progeny per male, x is the number of *SD* offspring and p is the mean segregation ratio for that cross (i.e. mean k). The parameters of the distribution are p and n. In the data we wish to analyse n is variable, thus one must calculate the binomial distribution with variable n. This was done by calculating $f(x)$ for each n observed and then weighting that by the frequency of n in that experiment. The number of *SD* progeny for each family size were then converted into k values that were grouped into histograms, according to their calculated frequencies. Parameter estimation was not a problem in this simulation, since mean k could be used for p in all families.

The important point to note here is that p was held constant at the mean k value of the particular cross being simulated. As a result the observed frequency distribution of k could be compared to a frequency distribution predicted assuming that the only source of variation in the segregation ratio was due to binomial sampling, with the added complication of varying family sizes.

The other model that was considered was one where binomial sampling was superimposed on an underlying distribution of the segregation ratio. We have defined k as the observed segregation ratio. Variation of k must consist in part, or entirely, of binomial sampling. We will define j as the potential segregation ratio of a male prior to binomial sampling. It follows from this definition that if the distribution of k conformed exactly to a binomial distribution, then the variance of j would be zero. The model we will consider next is one where j varies from male to male and the frequency distribution describing that variation is a beta distribution. This model predicts that k will be distributed as a beta-binomial variable.

Skellam (1948) described an explicit expression for a beta-binomial frequency distribution. This expression is:

$$f(x) = \binom{n}{x} \frac{B(\alpha+x, \beta+n-x)}{B(\alpha,\beta)},$$

where

$$B(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

and $\Gamma(x)$ is the generalized factorial of $(x+1)$ for non-integer x. This expression permits one to calculate the frequency of a particular class, x, when the family size is n and the parameters of the prior beta distribution are α and β.

A beta distribution has a mean of

$$\mu = \frac{\alpha}{\alpha+\beta},$$

and the variance is

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The mean of a beta distribution will be the same as the mean of a beta-binomial distribution, since the superposition of binomial sampling on a prior beta distribution will only alter the variance of the compound distribution. Thus, the mean of the observed k distribution can be used to determine the ratio of α to β, but the variance of the prior beta distribution (i.e. the variance of j) must be estimated in order to solve for exact values of α and β. A maximum likelihood method was used to estimate the variance of j. This was done by varying α, while keeping the ratio of α to $(\alpha+\beta)$ equal to the observed mean k, until the summed log likelihoods for all x and n observed were at a maximum. The probabilities were calculated using the formula for the beta-binomial distribution given above. The contribution of each n class to the total likelihood was weighted by the frequency of that class.

Once α and β had been estimated, one could then proceed to calculate the expected frequency distribution of k for a given experiment. This was done by substituting the estimated values of α and β into the formula for a beta-binomial distribution and then calculating the weighted frequency of each x class for a given n, where the weight was determined by the frequency of the n class.

The statistical assessment of differences between observed and predicted frequency distributions of k were determined by means of χ^2 tests. The frequency distributions consisted of intervals of 0·01, except where expected frequencies were less than five. Intervals with expected frequencies of less than five were pooled with adjacent intervals, until the expected frequency was greater than five. In the case of the comparisons with a beta-binomial distribution the degrees of freedom for the χ^2 test were three less than the number of classes in the resulting histograms. For comparisons with the predictions of the binomial distribution the degrees of freedom were two less than the number of classes.

3. Results and discussion

The observed k distributions were first compared to k distributions predicted if the only sources of variation were (1) binomial sampling from populations with the mean k value and (2) numbers of progeny per vial as observed in each experiment. (These predicted distributions were calculated assuming that the variance of j was zero.) The results of this analysis are summarized

https://doi.org/10.1017/S0016672300025489 Published online by Cambridge University Press

Table 1. *Comparison of observed and predicted distributions*

Chromosome	$\bar{k}$	s_k*	Binomial $\chi^2_{[d.f.]}$	Beta-binomial α	β	s_j†	Beta-binomial $\chi^2_{[d.f.]}$
SD-5	0·985	0·0223	$332_{[7]}$ $P < 10^{-6}$	60·6	0·921	0·0154	$7·33_{[8]}$ $P = 0·501$
R(SD-5)pk cn	0·932	0·0576	$1732_{[15]}$ $P < 10^{-6}$	22·8	1·67	0·0500	$43·0_{[23]}$ $P = 0·0069$
Df(2L)SD-5-8	0·878	0·0684	$1169_{[20]}$ $P < 10^{-6}$	32·5	4·50	0·0530	$38·0_{[29]}$ $P = 0·124$
RR(SD-5)lt	0·713	0·1175	$3606_{[29]}$ $P < 10^{-6}$	14·3	5·75	0·0986	$62·2_{[48]}$ $P = 0·082$
R(SD-5)b pr-5	0·506	0·0644	$43·6_{[27]}$ $P = 0·023$	204	200	0·0248	$27·6_{[28]}$ $P = 0·485$
RR(SD-5)pr lt	0·463	0·0760	$92·8_{[33]}$ $P < 10^{-6}$	82·6	96·0	0·0372	$45·1_{[34]}$ $P = 0·0966$

* s_k refers to the observed standard deviation of the segregation ratio, k (which is expected to be distributed as a beta-binomial distribution).
† s_j refers to the standard deviation of j (which is defined as the prior distribution of k and will be distributed as a beta distribution, without the effects of binomial sampling).

in column 4 of Table 1. It is apparent that, at the 0·05 level of significance, none of the *SD* chromosomes that were tested showed agreement with the k distribution predicted assuming that the only sources of variation were binomial sampling and family size.

These results demonstrate rigorously that the variation of segregation ratios of *SD* males does not conform to a binomial distribution with variable n. The most likely explanation for this observation is that the potential segregation ratio of males prior to binomial sampling, j, is not constant from male to male. If j did vary from male to male, then binomial sampling would be superimposed upon the prior distribution of j, due to the sampling of gametes from the total gamete pool of individual males. One possible model to describe such variation is a beta-binomial distribution (Skellam, 1948). The results of fitting the six *SD* distributions to a beta-binomial distribution are shown in the last four columns of Table 1. One can see that all fit a beta-binomial distribution with a probability of greater than 0·05, except *R(SD-5)pk cn*, which only fits with a probability of 0·0069.

The shapes of the observed and predicted distributions are shown in Fig. 1. It is apparent that, except for *R(SD-5)pk cn*, the observed and predicted distributions parallel each other quite closely, with no obvious systematic deviations. The histograms shown for *R(SD-5)pk cn*, however, do show a marked deficiency of predicted values in the 0·83–0·91 interval and an excess of predicted values in the 0·92–0·98 interval. The statistical significance of this systematic pattern of deviations is confirmed by the χ^2 value of 43·0 (with 23 d.f.). Close examination of the observed distribution of *R(SD-5)pk cn* shows that there is a shoulder in the 0·83 to 0·91 interval. This strongly indicates that the observed *R(SD-5)pk cn* distribution was the result of the fusion of two distributions with slightly different means. This could be caused by the presence of a fairly major modifier of *SD* present at a frequency near 0·5 in the population of *R(SD-5)pk cn/cn bw* males used to generate the observed distribution. Thus, lack of fit of the observed distribution to a beta-binomial distribution could be the result of trying to fit a bimodal observed distribution to a unimodal predicted distribution.

The data, in general, demonstrate that the variation of *SD* segregation ratios can be adequately described by a beta-binomial distribution as is the case for sex ratios in humans. Edwards (1958) analysed the data of Geissler (1889) on sex ratios of 991 958 families. The sex ratios showed significant variation in excess of that predicted by the binomial distribution alone and the excess variation could be explained by a beta distribution. The estimate for the standard deviation of the prior beta distribution of sex ratios was 0·05. This is even larger than the estimates of the standard deviation of j shown in Table 1 for the k distributions with means near 0·5, i.e. *R(SD-5)b pr-5* and *RR(SD-5)pr lt* (0·0248 and 0·0372 respectively). However, not all human populations show evidence for heterogeneity of sex ratios. Edwards & Fraccaro (1958) analysed the sex ratios of families from 5477 Swedish ministers and they could not find any evidence for variation in excess of binomial sampling. Thus, although human populations can show variation in segregation ratios similar to that shown here with *SD* chromosomes, not all populations demonstrate the same degree of heterogeneity.

This observation leads one to consider the nature of the heterogeneity. As for any type of phenotypic variation, it could consist of both genetic and environmental components. One common method

https://doi.org/10.1017/S0016672300025489 Published online by Cambridge University Press

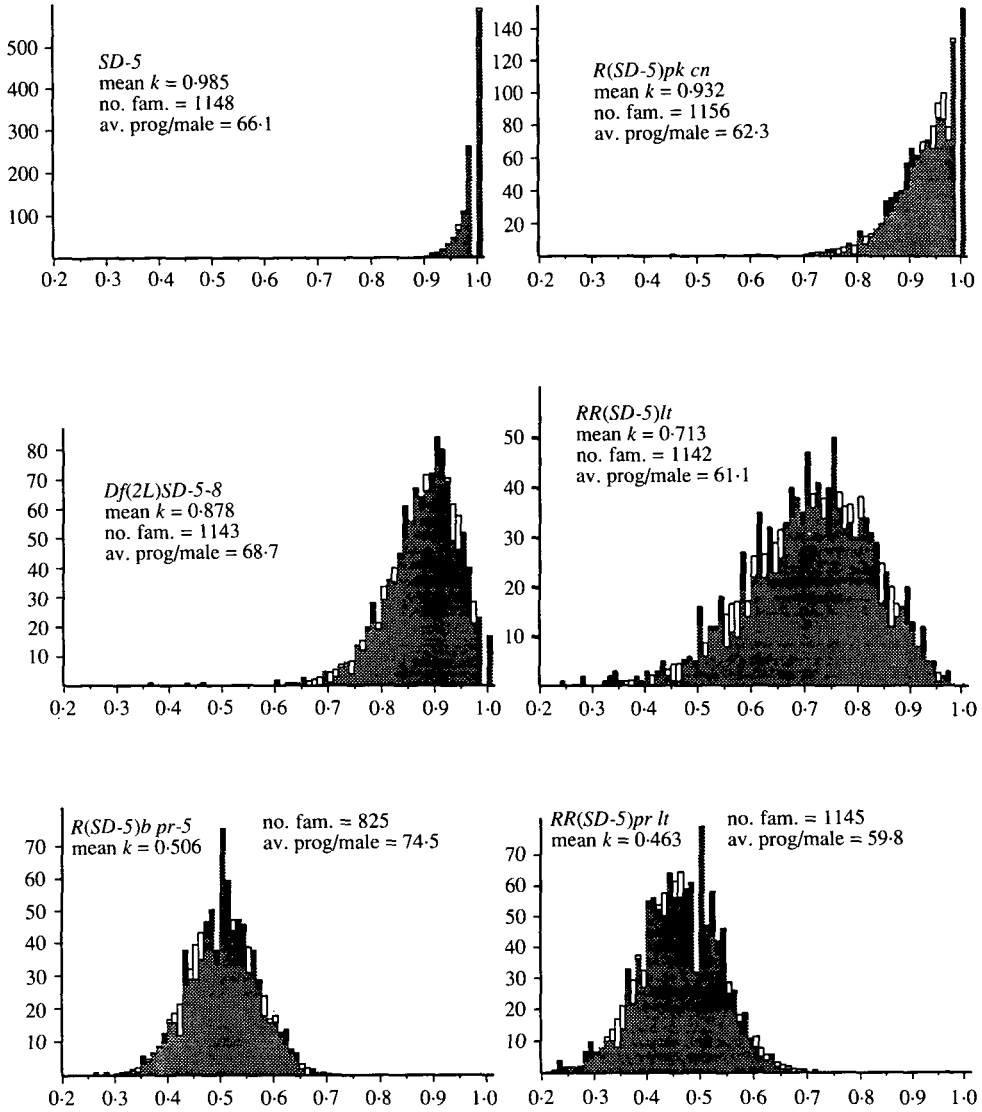

Fig. 1. Observed distributions compared with predictions of the beta-binomial distribution. The ordinate in all graphs is the number of families in that interval, while the abscissa is k, in intervals of 0·01. Regions where predicted and observed distributions overlap are indicated by stipling. The solid black regions indicate where observed frequencies exceed predicted frequencies and the open rectangles represent an excess of predicted over observed values. Each graph is also labelled with: the chromosome used, the unweighted mean k for that experiment, the total number of families scored and the average number of progeny per male for that experiment. The paucity of observations in the interval $0{\cdot}99 \leqslant k < 1{\cdot}0$ is due to the small number of families with more than 100 offspring.

that is used to partition variation into genetic and environmental components is to use parent offspring regression. In the case of *SD* this would be virtually impossible to do, because the genetic modifiers of *SD* that one wants to detect do not generally segregate independently from the *SD* chromosome, which must be directly selected each generation. (Recall that there is no recombination in *D. melanogaster* males and that there are only two large autosomes and the dot fourth chromosome.) Another possible method that could be used for the biometrical analysis of *SD* would be to compare the amount of variation of inbred and outbred lines. The work described here provides a framework for this type of analysis, as the variation one would want to examine would be the variance of j, not the variance of k. Since most *SD* lines show a variance of k which conforms reasonably well to a beta-binomial distribution, the prior beta distribution should provide a reasonable approximation for the statistical description of the variation of j. By using maximum likelihood methods as described herein, one could estimate α and β, and therefore the variance of j, for both inbred and outbred *SD* lines. Without this methodology it would be virtually impossible to obtain a biometrical description of segregation distortion. Similar methods could be applied to the analysis of virtually any segregation ratio, such as sex ratios or that of any pair of alleles.

This research was supported by Natural Sciences and Engineering Research Council of Canada grants 851-78 to A. J. Hilliker and A5853 to D. G. Holm. C. B. Sharp was the recipient of a NSERC postgraduate scholarship.

References

Brittnacher, J. G. & Ganetzky, B. (1983). On the components of segregation distortion in *Drosophila melanogaster*. II. Deletion mapping and dosage analysis of the *SD* locus. *Genetics* **103**, 659–673.

Brittnacher, J. G. & Ganetzky, B. (1984). On the components of segregation distortion in *Drosophila melanogaster*. III. Nature of enhancer of *SD*. *Genetics* **107**, 423–434.

Edwards, A. W. F. (1958). An analysis of Geissler's data on the human sex ratio. *Annals of Human Genetics* **23**, 6–15.

Edwards, A. W. F. & Fraccaro, M. (1958). The sex distribution in the offspring of 5477 Swedish ministers of religion, 1585–1920. *Hereditas* **44**, 447–450.

Ganetzky, B. (1977). On the components of segregation distortion in *Drosophila melanogaster*. *Genetics* **86**, 321–355.

Geissler, A. (1889). Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. *Z. K. Sächs. Stat. Bureaus* **35**, 1–24.

Hartl, D. L. (1974). Genetic dissection of segregation distortion. I. Suicide combinations of *SD* genes. *Genetics* **76**, 477–486.

Hiraizumi, Y. & Crow, J. F. (1960). Heterozygous effects on viability, fertility, rate of development and longevity of *Drosophila* chromosomes that are lethal when homozygous. *Genetics* **45**, 1071–1084.

Hiraizumi, Y., Martin, D. W. & Eckstrand, I. A. (1980). A modified model of segregation distortion in *Drosophila melanogaster*. *Genetics* **95**, 693–706.

Lyttle, T. W., Brittnacher, J. G. & Ganetzky, B. (1986). Detection of *Rsp* and modifier variation in the meiotic drive system *Segregation Distorter* (*SD*) of *Drosophila melanogaster*. *Genetics* **114**, 183–202.

Martin, D. W. & Hiraizumi, Y. (1979). On the models of segregation distortion in *Drosophila melanogaster*. *Genetics* **93**, 423–435.

Miklos, G. L. G. & Smith-White, S. (1971). An analysis of the instability of segregation-distorter in *Drosophila melanogaster*. *Genetics* **67**, 305–317.

Nicoletti, B. (1968). Il controllo genetico della meiosi. *Atti. Associazione Genetica Italiana* **13**, 1–71.

Sandler, L. & Hiraizumi, Y. (1960). Meiotic drive in natural populations of *Drosophila melanogaster*. IV. Instability at the segregation-distorter locus. *Genetics* **45**, 1269–1287.

Sandler, L. & Hiraizumi, Y. (1961). Meiotic drive in natural populations of *Drosophila melanogaster*. VIII. A heritable ageing effect on the phenomenon of segregation distortion. *Canadian Journal of Genetics and Cytology* **3**, 34–36.

Sandler, L., Hiraizumi, Y. & Sandler, I. (1959). Meiotic drive in natural populations of *Drosophila melanogaster*. I. The cytogenetic basis of segregation-distortion. *Genetics* **44**, 233–250.

Sharp, C. B. & Hilliker, A. J. (1986). Measuring segregation distortion in *Drosophila melanogaster*. *Canadian Journal of Genetics and Cytology* **28**, 395–400.

Sharp, C. B. & Hilliker, A. J. (1987). The *k*-probit transformation and segregation distortion. *Drosophila Information Service* **66**, 127–128.

Sharp, C. B., Hilliker, A. J. & Holm, D. G. (1985). Further characterization of genetic elements associated with the segregation distortion phenomenon in *Drosophila melanogaster*. *Genetics* **110**, 671–688.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society* B **10**, 257–261.

Temin, R. G. & Marthas, M. (1984). Factors influencing the effect of segregation distortion in natural populations of *Drosophila melanogaster*. *Genetics* **107**, 375–393.

Tokuyasu, K. T., Peacock, W. J. & Hardy, R. W. (1977). Dynamics of spermiogenesis in *Drosophila melanogaster*. VII. Effects of segregation distorter (*SD*) chromosome. *Journal of Ultrastructure Research* **58**, 96–107.

Wu, C.-I., Lyttle, T. W., Wu, M. L. & Lin, G. F. (1988). Association between a satellite DNA sequence and the *Responder* of segregation distortion in *D. melanogaster*. *Cell* **54**, 179–189.