Genet. Res., Camb. (1996), **68**, *pp.* 131–149 *With 7 text-figures*

Background selection and patterns of genetic diversity in *Drosophila melanogaster*

BRIAN CHARLESWORTH
Department of Ecology and Evolution, The University of Chicago, 1101 E. 57th Street, Chicago, IL 60637-1573, USA

(*Received 11 September 1995 and in revised form 26 February 1996*)

Summary

Theoretical models of the effects of selection against deleterious mutations on variation at linked neutral sites (*background selection*) are used to predict the relations between chromosomal location and genetic variability at the DNA level, in *Drosophila melanogaster*. The sensitivity of the predictions to variation in the mutation, selection and recombination parameters on which they are based is examined. It is shown that many features of the observed relations between chromosomal location and level of genetic diversity in *D. melanogaster* can be explained by background selection, especially if the weak selective forces acting on transposable elements are taken into account. In particular, the gradient in diversity in the distal portion of the X chromosome, and the lack of diversity on chromosome 4 and at the bases of the major chromosomes, can be fully accounted for. There are, however, discrepancies between predicted and observed values for some loci in *D. melanogaster*, which may reflect the effects of forces other than background selection.

1. Introduction

Several regions of the genome of *Drosophila melanogaster* exhibit greatly reduced rates of meiotic crossing over. These include the telomeric region of the X chromosome, the pericentric regions of the major chromosomes, and chromosome four (which lacks meiotic exchange under normal conditions) (Ashburner, 1989, chap. ii). Surveys of DNA variation in natural populations of *D. melanogaster* strongly suggest that genetic variability is lower in such regions, relative to its level in regions where crossing over occurs at normal frequencies (Aguadé *et al.* 1989*a*, *b*; Aguadé & Langley, 1994; Moriyama & Powell, 1996). As shown by Begun & Aquadro (1992) and Aquadro *et al.* (1994), there is a close relationship between the frequency of recombination per nucleotide site in the neighbourhood of a gene, and its level of within-population variability.

This result can be interpreted in at least two ways. Hitch-hiking of neutral or nearly neutral variants by more strongly selected favourable mutations at closely linked loci can cause a substantial loss in variability (Maynard Smith & Haigh, 1974; Kaplan *et al.* 1989). Linkage of such variants to deleterious mutant alleles, which are destined to be rapidly eliminated from the population, also reduces variability (Charlesworth *et al.* 1993, 1995; Charlesworth, 1994; Hudson, 1994; Hudson & Kaplan, 1994, 1995). These two processes are often referred to as 'selective sweeps' (Berry *et al.* 1991) and 'background selection' (Charlesworth *et al.* 1993), respectively. Temporal fluctuations in the direction of selection on segregating alleles may also affect variation at linked sites (Gillespie, 1994; Barton, 1995).

One way of evaluating the relative contributions of these processes to the patterns displayed in the data is to develop theoretical models of the relations between the recombinational environment and its level of neutral genetic variability. Several such models have been developed for the case of selective sweeps (Wiehe & Stephan, 1993; Braverman *et al.*, 1995; Simonsen *et al.* 1995; Stephan, 1995). Hudson & Kaplan (1995) have modelled background selection arising from multiple loci subject to mutation and selection, using the simplifying assumptions of at most one crossover per chromosome and equal selective effects at each locus. They have used their results to predict the pattern of genetic variability as a function of chromosomal location within the genome of *D. melanogaster*.

The purpose of this paper is to develop models of the effects of background selection on neutral variation in *D. melanogaster*, using the general theoretical framework developed by Hudson & Kaplan (1995)

and Nordborg *et al.* (1996). In particular, I allow for variation among mutant alleles in their effects on fitness, specifically including the weak selection coefficients against transposable element (TE) insertions indicated by population survey results (Charlesworth *et al.* 1992*a*). While weakly selected deleterious alleles have little effect on diversity in freely recombining regions, they can have a substantial effect when there is little or no recombination (Nordborg *et al.* 1996), so that it is important to incorporate them into the models. The results suggest that many features of the patterning of genetic variation across different chromosomal regions in *D. melanogaster* can be predicted by background selection, although some aspects of the data seem to require explanation in terms of other factors.

2. Construction of models of *Drosophila* chromosomes

(i) *Effect of detrimental mutations on a linked neutral site*

Theoretical and simulation results (Hudson & Kaplan, 1995; Nordborg *et al.* 1996) suggest that, for a large random-mating population, the effect of background selection on a neutral site embedded in a set of linked autosomal loci subject to mutation to deleterious alleles is predicted by the formula

$$\frac{\pi}{\pi_0} \approx \exp -\sum_i \frac{q_i}{(1+\rho_i)^2}, \quad (1)$$

where π/π_0 is the ratio of the nucleotide site diversity at the site in question to the classical neutral value, and q_i is the equilibrium mutant allele frequency at the ith selected locus. If the mutation rate at the ith locus is u_i, $q_i = u_i/t_i$, where t_i is the heterozygous fitness effect of the mutant allele at that locus (it is assumed that t_i is $\gg u_i$). Also, $\rho_i = r_i(1-t_i)/t_i$, where r_i is the recombination frequency between the neutral locus and the ith selected locus (Nordborg *et al.* 1996). Note that the effective autosomal recombination frequency in *Drosophila* is half the recombination frequency in female meiosis, due to the lack of crossing over in males.

This formula appears to be quite robust to the details of the mode of selection, and gives good fits to simulation results, provided that population sizes are of the order of a few thousand individuals and selection is not too weak [see Hudson & Kaplan (1995), Nordborg *et al.* (1996), and Section 4]. It can be used to predict the amount of variation at neutral sites anywhere in the genome, provided that we know the mutation, selection and recombination parameters involved.

Since mutation rates and fitness effects at individual locus are unknown, the models are based on estimates of mutation and selection parameters for the whole second chromosome of *D. melanogaster*, obtained from the mutation accumulation experiments of Mukai *et al.* 1972) and Ohnishi (1977) on detrimental (minor-effect) mutations on the second chromosome which affect egg-to-adult viability. The small contribution from lethals to the overall mutation rate to deleterious alleles and to the equilibrium mean number of deleterious mutations per genome (Crow & Simmons, 1983) will be ignored. These experiments have recently been re-analysed by Keightley (1994), who showed that there is evidence for substantial variation in selection coefficients among new mutations; if this is taken into account, the mutation rate per chromosome is much greater than that obtained with the assumption of equal mutation rates (Crow & Simmons, 1983; Keightley, 1994).

Unfortunately, the data do not provide a satisfactory joint estimate of the mutation rate and the parameters of the distribution of selection coefficients (Keightley, 1994). Somewhat conservatively (Keightley, 1994), I assume a mutation rate of about 0·10 for the X chromosome, 0·18 for chromosome 2 and 0·2 for chromosome 3, in accordance with the approximate relative sizes of the euchromatic sections of these chromosomes (Charlesworth *et al.* 1992*b*; Heino *et al.* 1994). There is evidence for strong positive correlations between the effects of detrimental mutations on different fitness components (Houle *et al.* 1994). This implies that the net effect of a detrimental mutation on fitness is likely to be much larger than its effect on viability. I assume that the homozygous effect of a mutation on net fitness is 4 times its effects on viability.

I also assume that homozygous mutational effects on viability are distributed according to a gamma distribution with parameters $\alpha = 31{\cdot}4$ and $\beta = 0{\cdot}691$, the maximum likelihood estimates from the data of Mukai *et al.* (1972) when a second chromosome mutation rate of 0·2 is assumed (Keightley, 1994, table 2). Ohnishi's (1977) data yield a smaller mean selection coefficient and a larger coefficient of variation (Keightley, 1994). As shown in Section 3, the use of a larger mean and a narrower spread of selection coefficients yields a smaller overall effect of background selection, and a lesser sensitivity to variation in recombination rates. The values used here, therefore, predict a smaller overall effect of background selection and a looser relation between location and genetic diversity than would be obtained with the parameters estimated from Ohnishi's data for the same mutation rate. In this sense, they are conservative.

Assuming a dominance coefficient of 0·2, as suggested by data on the effects on several fitness components of mutations segregating in natural populations (Crow & Simmons, 1983; Hughes, 1995), the above parameters yield a mean heterozygous effect of mutations on fitness of 0·018. This lies between the value of 0·02 of the harmonic mean of the t_i with respect to net fitness obtained by Crow & Simmons (1983), and the value of 0·01 obtained by Charlesworth

& Hughes (unpublished). Since very weakly selected loci do not contribute significantly to background selection (Charlesworth *et al.*, 1993; Nordborg *et al.* 1996; Section 4), I truncate the gamma distribution below a homozygous viability selection coefficient of 10^{-4} when applying eqn (2) below. In practice, truncation has little effect, since less than 2% of all mutations fall into this category with the values of α and β used here.

Loci are assumed to be distributed uniformly along the physical map of the chromosome in question. The physical position of a locus subject to mutation and selection is represented by the variable z ($0 \leqslant z \leqslant 1$), which is the proportion of the length of the chromosome from an end that is arbitrarily taken to be the origin. The frequency of recombination between a selected locus at position z and the neutral site under consideration is denoted by $r(z)$; mutant loci in the neighbourhood of position z are assumed to follow the distribution of selective effects described above, with probability density $\phi(t)$ for selection coefficient t. Mutation rates are assumed to be independent of location and selection coefficient, and the sum of the haploid mutation rates per locus over the chromosome is u. The sum in eqn (1) is thus approximated by the integral

$$\frac{\pi}{\pi_0} \approx \exp -u \int_0^1 \int_{t_0}^1 \frac{\phi(t)}{t(1+\rho[t,z])^2} \mathrm{d}t\, dz, \qquad (2)$$

where $\rho(t,z) = r(z)(1-t)/t$, and t_0 is the minimum value of t considered as effective in contributing to background selection (Nordborg *et al.* 1996).

Some modification is needed for the case of X-linked mutations, which are selected against in the hemizygous state in males. The most conservative assumption is that all deleterious mutations act similarly in both males and females, although many are in fact sex-limited in their effects on fitness (Ashburner, 1989, chap. 10). The equilibrium frequency of a mutant allele at locus i is then $q_i = u_i/\tilde{t}_i$, where $\tilde{t}_i = (2t_i + s_i)/3$, and s_i is its homozygous or hemizygous effect on fitness (Haldane, 1927). This can be substituted into eqns (1) and (2), replacing t_i by $\tilde{t}_i$. The expression for ρ_i is unchanged, except that the effective recombination rate for X chromosomal loci is two-thirds the value for female meiosis, since recombination on the X is effective only in females. Since selection is stronger against rare deleterious alleles at sex-linked loci than at autosomal loci with equivalent effects on fitness, I use a truncation point of 10^{-5} for the homozygous viability effects of X-linked mutations.

(ii) *Effect of transposable elements on a linked neutral site*

Transposable elements may be a source of background selection which is not included in the estimates of mutational parameters discussed above (Hudson, 1994). Deleterious chromosomal rearrangements due to meiotic ectopic exchange between TEs seem to play an important role in the elimination from the population of elements which insert into genomic regions where they do little or no direct mutational damage (Langley *et al.* 1988; Montgomery *et al.* 1991). Since the mutation accumulation experiments described above use chromosomes transmitted exclusively through males, where meiotic recombination is absent, no such deleterious consequences of meiotic ectopic exchange will be detected in these experiments, although any deleterious effects of insertional mutations or mitotic ectopic exchange would contribute to the estimated mutation rates.

Statistical analyses of data on element distribution in natural populations of *D. melanogaster* indicate that the selection coefficients against naturally occurring heterozygous element insertions are very small, of the order of 2×10^{-4} on average, but deterministic forces seem to be sufficiently large in relation to the effects of genetic drift that TEs are held at low frequencies in most parts of the genome (Charlesworth *et al.* 1992*a*). Selection of this magnitude is consistent with the evidence that ectopic exchange involving a given TE insertion is a low frequency event (Montgomery *et al.* 1991). The fraction of individuals carrying a TE insertion at a given site which are eliminated by selection is thus very small, even if rearrangements with a severe effect on fitness are generated by ectopic exchange. The frequency of a TE insertion at a given chromosomal site thus behaves formally like that of a weakly selected, deleterious allele, so that eqn (1) can be applied. Although the average number of new TE insertions per generation is at least 0·1 per haploid genome when summed over all families (Nuzhdin & Mackay, 1995), a mean selection coefficient of 2×10^{-4} means that they could contribute only a small part of the observed rate of mutational decline in viability (Crow & Simmons, 1983; Keightley, 1994), even if ectopic exchange is not their primary mode of elimination. This provides an additional reason for treating TEs separately.

It is assumed that TEs are in equilibrium between transposition and selection, with selection coefficient t' against a heterozygous element insertion. If elements are in approximate equilibrium between the effects of transposition and removal by ectopic exchange and other forces, as is indicated by the population data, t' should be the same as the rate of transposition per element per generation, regardless of genomic location and the associated rate of ectopic exchange (Charlesworth, 1991). Published estimates of mean transposition rates in the literature (Nuzhdin & Mackay, 1995) agree approximately with the above estimate of the rate of elimination, supporting this assumption.

If we pool over all element families and treat the chromosome as a continuum, the equivalent of q_i in eqn (1) is $n(z)\mathrm{d}z$, the mean number of elements per

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

chromosome found between locations z and $z+\mathrm{d}z$. If different element families have different rates of elimination, with probability density distribution $\phi(t')$, eqns (1) and (2) can be modified to yield the following expression for the contribution of TEs to the effect of background selection on variability at a given neutral site:

$$\exp - \int_0^1 n(z) \int_{t'_0}^1 \frac{\phi(t')}{(1+\rho[t', z])^2} \mathrm{d}t' \, \mathrm{d}z, \qquad (3a)$$

where $\rho(t', z) = r(z)/t'$ (the weak selection here means that the inclusion of the term $(1-t')$ in the numerator is unnecessary).

If the variance in t' among element families, $V_{t'}$, is not too large, this expression can be approximated by expanding around the harmonic mean of t', $\tilde{t}$, which yields the expression

$$\exp - \int_0^1 \frac{n(z)}{(1+\rho[\tilde{t}, z])^2} \left\{ 1 + \frac{3 V_{t2} \, r(z)^2}{\tilde{t}^4 (1+\rho[\tilde{t}, z])^2} \right\} \mathrm{d}z. \qquad (3b)$$

This result shows that a model that includes variance in transposition/elimination rates among families has a larger effect on genetic diversity than if rates are constant. The use of the above estimate of the harmonic mean rate of transposition/elimination, but ignoring variance in the rate, should therefore provide a conservative estimate of the contribution of TEs to background selection.

The contribution from TEs was estimated from data on the ten families of elements surveyed by *in situ* hybridization of element probes to chromosomes sampled from a Maryland population by Charlesworth & Lapid (1989) and Charlesworth *et al.* (1992*a*), with additional information from unpublished data on a further six families surveyed on 10 X chromosomes from the same population (Charlesworth, Assimacopoulos & Britton, unpublished). Given the evidence for an accumulation of elements at the bases of the chromosomes, the chromosome arms are divided into proximal and distal portions for the purpose of calculating element densities, combining the tip and mid sections as defined by Langley *et al.* (1988) to form the distal portion of each arm. The term $n(z)$ in eqn (3*a*) is assumed to be independent of z within each portion of the chromosome, and is thus equal to the mean number of elements per chromosome for that portion of the chromosome, divided by the proportion of the chromosome which it represents.

The distal portion of the X comprises 90% of the X chromosome as defined in Section (iii) below. Overall, the mean copy numbers of the 16 families of TEs are 39·0 (distal) and 17·3 (proximal). There are 40–50 families of TEs in *D. melanogaster* (Lindsley & Zimm, 1992), so that this represents the count for approximately one-third of all element families. More abundant families were more likely to be cloned in the early investigations of the *D. melanogaster* genome than less abundant ones, so that the elements not included in these surveys may well form less than two-thirds of the total number per genome. A conservative estimate of the total number of elements per chromosome is probably given by multiplying the above values by 1·5, giving a final estimate of 58·5 and 25·9 for the distal and proximal portions of the X, respectively. An alternative approach is to use estimates of element density per kilobase obtained from population surveys of restriction fragment length polymorphisms (RFLPS) in small genomic regions. The estimate of a mean element density of elements per kilobase of 0·005 for the euchromatin (Charlesworth & Langley, 1991) implies a total of about 110 elements for the X chromosomal euchromatin, if elements are randomly distributed among chromosomes. This is substantially larger than the *in situ* estimate, but should probably be reduced by about 15% because of the under-representation of elements on the X chromosome relative to the autosomes (Biémont, 1992; Charlesworth *et al.* 1992*b*). The *in situ* estimate thus appears to be a slightly conservative value for the abundances of TEs, and is used in what follows.

In the case of the autosomes, the *in situ* estimates of element abundances of Charlesworth *et al.* (1992*b*) are simply multiplied by a factor of 2, to account for the families which have not been surveyed. The same procedure applied to the X gives slightly smaller values than those obtained above, so that this is a conservative procedure, as is also indicated by the calculations of autosomal element densities from the RFLP studies. For 2L, the values are 58·1 and 15·2 for the distal and proximal portions (which contribute 41·2% and 7·2% of chromosome 2, respectively). For 2R, the values are 58·5 and 15·0 for the distal and proximal portions (44·7% and 6·9% of chromosome 2). For 3L, the corresponding numbers are 54·4 and 18·0 (38·1% and 6·15% of chromosome 3), and for 3R they are 84·4 and 22·2 (47·0% and 8·75% of chromosome 3).

(iii) Drosophila *chromosome models: general considerations*

For chromosomal regions consisting of many polytene chromosome bands, there is a good approximate correspondence between the amount of DNA in the regions concerned and the number of bands visible in the Lefevre (1976) photographic map (Montgomery *et al.* 1987; Charlesworth & Lapid, 1989; Charlesworth *et al.* 1992*b*). I have therefore used the band coding system of Charlesworth & Lapid (1989) and Charlesworth *et al.* (1992*a*), based on the enumeration of these polytene bands, to assign approximate physical positions to loci for which information on DNA variability is available. For some extreme proximal locations, the numbering system has been extended slightly, as described below. Such band

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

counts may well be inaccurate indicators of the physical size of small regions; in addition, accurate information on genetic map distances for such regions is often unavailable. There is also abundant evidence in the literature for considerable genetic and environmental variation in recombination frequencies for a given interval, and the standard genetic map distances cannot be used as accurate indicators of the actual amount of crossing over in small regions. This is especially true for centromere-proximal regions, which are highly sensitive to such variation and to the inter-chromosomal effect of inversion heterozygosity on crossing over (Lucchesi, 1976; Brooks, 1988; Sniegowski *et al.* 1994).

For these reasons, I have relied as far as possible on published data on rates of crossing over for defined intervals near telomeres and centromeres, where exchange rates per nucleotide are lowest (Ashburner, 1989, chap. 11), and where predictions of the effects of background selection are likely to be most sensitive to the recombination parameters used in the models. Since natural populations are polymorphic for inversions, the inter-chromosomal effect will cause effective rates of crossing over to be higher than those in crosses with homokaryotypic backgrounds, possibly by as much as 50% or more for centromeric or telomeric regions (Lucchesi, 1976; Sniegowski *et al.* 1994). Conversely, suppression of exchange in inversion heterozygotes means that standard maps may overestimate the effective frequency of recombination in the mid-arm sections of the autosomes, where most naturally occurring inversions are located (Lemeunier & Aulard, 1992). To provide conservative estimates of the effects of background selection, I have generally used high rather than low estimates of rates of proximal and distal exchange whenever alternative values are available (see Appendix, Sections (i)–(iii)).

The calculation of the effect of background selection on a neutral gene from eqns (2) and (3) requires knowledge of its frequencies of recombination with all selected loci on the chromosome on which it is located. (The very weak effect of background selection from deleterious alleles on a different chromosome (Nordborg *et al.* 1996) means that it is legitimate to consider only a single chromosome.) If the physical distance between the marker locus and a selected locus is z, knowledge of the recombination parameters for the interval separating the two loci enables calculation of the map distance between them, $l(z)$, as described below. To determine the corresponding recombination frequency, $r(l[z])$, a mapping function relating r and l is needed. Cobbs (1978) has shown that the following mapping function, suggested by Owen (1951), appears to fit the *Drosophila* data for a single chromosome arm:

$$r(l[z]) = 0{\cdot}5(1-\cos(2l[z])\,e^{-2l[z]}). \quad (4)$$

This will be used in what follows for describing recombination within a chromosome arm, except where the effects of alternative mapping functions are explicitly mentioned.

For the metacentric chromosomes 2 and 3, there is strong negative interference across the centromere between intervals which are very close to the centromere, whereas there is no such effect between more distal regions (Ashburner, 1989, p. 470). Thus, although an exchange near the centromeric heterochromatin is a rare event, the conditional probability of a proximal exchange in one arm of an autosome may be increased by a factor of 10 or more by the occurrence of a proximal exchange in the other arm (Green, 1975; Sinclair, 1975; Denell & Keppy, 1979). The above mapping function thus cannot be used for determining the frequency of recombination between an autosomal marker locus and a selected locus on a different arm of the same chromosome. Special formulae, derived in the Appendix, Section (iv), must be used for this purpose.

Given the numerous sources of uncertainty in the data, there seems little point in trying to establish a fine-scale relation between physical position and recombination parameters, as has been done previously by the use of coefficients of exchange derived from the standard maps (Lindsley & Sandler, 1977; Begun & Aquadro; 1992; Wiehe & Stephan, 1993; Aquadro *et al.* 1994; Hudson & Kaplan, 1995; Stephan, 1995; Moriyama & Powell, 1996). Rather, I have divided each chromosome into several regions, and attempted to provide a rough description of the relation between physical location and genetic map for each region.

(iv) *The X chromosome*

The X chromosome is considered here to comprise the polytene map from 1A1 to 20F, inclusive. Charlesworth & Lapid (1989) assigned 172 bands to the euchromatic section 1A1 to 20A; there are seven more bands visible in Lefevre's (1976) photographic map in the region between 20B and 20F, bringing the total X chromosome count to 179. Since the genetic and molecular evidence suggests a nearly normal density of genes in this region (Schalet & Lefevre, 1976; Miklos & Cotsell 1990), despite its heterochromatic appearance, it seems appropriate to include it in the count of mutable sites. The physical locations of the proximal termini of the regions into which the X has been divided are denoted by $a_{1i}(i = 1\text{–}6)$, where a_{1i} is the proportion of the total number of X chromosome bands from the tip to the band corresponding to the proximal boundary of the region. The physical location of a gene, z, is defined as the proportion of the total length of the X chromosome separating it from the telomere. A gene is assigned to region i if $a_{1(i-1)} \leqslant z < a_{1i}$. The upper panel of Fig. 1 is a schematic representation of the assumed relations between recombination rate per unit physical distance and location on the X chromosome, and the numerical

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

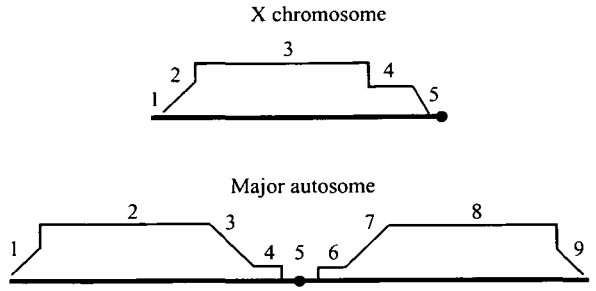

Fig. 1. The gradient of recombination frequency with respect to physical position on the chromosome, plotted against location, for the X chromosome and for the major autosomes (not to scale). The numbers denote the regions described in the Appendix, Sections (i)–(iii).

values of the parameters used to describe this chromosome are given in Table 1. The justifications for these are described in the Appendix, Section (i).

(v) *Chromosomes 2 and 3*

The euchromatic portion of chromosome 2 as defined here has a total of 311 bands (Table 2). 2L has been divided into regions whose proximal boundaries are denoted by a_{2i} ($i = 1$–4), where a_{2i} is the proportion of the total number of 2L chromosome bands from the tip to the band corresponding to the proximal boundary of the region. 2R is similarly divided into regions whose distal boundaries are denoted by a_{2i} ($i = 6$–9), where a_{2i} is the proportion of the total number of bands from the tip to 2L to the band corresponding to the distal boundary of the region in question. The fifth region of the chromosome is composed of the euchromatic portions of 2L and 2R which straddle the centromere, and its rightmost boundary is defined by a_{25}. The centromere is arbitrarily placed midway between a_{24} and a_{25}. The physical location of a gene, z, is denoted by the proportion of the total length of the chromosome 2 which separates it from the telomere of 2L. A gene is assigned to region i of 2L if $a_{2(i-1)} \leqslant z < a_{2i} (i \leqslant 4)$, to region i of 2R if $a_{2(i-1)} < z \leqslant a_{2i} (i \geqslant 6)$, and to region 5 if $a_{24} \leqslant z \leqslant a_{25}$. 2R is treated as a reflection of 2L, except for different numerical values of the coefficients in the functions describing recombination. A schematic representation of the relation between recombination rate per unit physical distance and location is shown in the lower panel of Fig. 1. The second chromosome parameters shown in Table 2 are justified in the Appendix, Section (ii).

Table 1. *Parameters of the X chromosome*

Region	Proximal boundary	Recombination function*	Polynomial coefficients
1	1B4, $a_{11} = 0{\cdot}017$ (3 bands)	Linear	$c_{11} = 0{\cdot}011$
2	3C2, $a_{12} = 0{\cdot}123$ (22 bands)	Quadratic	c_{11}, $b_{11} = 1{\cdot}21$
3	15F1–3, $a_{13} = 0{\cdot}799$ (143 bands)	Linear	$c_{12} = 0{\cdot}817$
4	19D3, $a_{14} = 0{\cdot}939$ (168 bands)	Linear	$c_{13} = 0{\cdot}640$
5	20C1, $a_{15} = 0{\cdot}989$ (177 bands)	Quadratic	c_{13}, $b_{12} = 6{\cdot}40$
6	20F, $a_{16} = 1{\cdot}00$ (179 bands)	No recombination	

The numbers of bands (in brackets) refer to the number from the telomere of the X, in the numbering system of Charlesworth *et al.* (1992*a*).
* This refers to the nature of the polynomial relating map distance to physical distance in the region in question. See Appendix for further details.

Table 2. *Parameters of chromosome 2*

Region	Proximal boundary (2L) or distal boundary (2R)	Recombination function	Polynomial coefficients
2L			
1	22A1, $a_{21} = 0{\cdot}026$ (8 bands)	Quadratic	$c_{21} = 0{\cdot}020$, $b_{21} = 29{\cdot}5$
2	31A1, $a_{22} = 0{\cdot}222$ (69 bands)	Linear	$c_{22} = 1{\cdot}94$
3	38A1, $a_{23} = 0{\cdot}412$ (128 bands)	Quadratic	$c_{23} = 0{\cdot}143$, $b_{22} = 4{\cdot}70$
4	40C1, $a_{24} = 0{\cdot}482$ (150 bands)	Linear	c_{23}
2L/2R (pericentric region)			
5	41E1, $a_{25} = 0{\cdot}486$ (151 bands)	No recombination	
2R			
6	42F3, $a_{26} = 0{\cdot}527$ (164 bands)	Linear	$c_{24} = 0{\cdot}171$
7	50F9, $a_{27} = 0{\cdot}736$ (229 bands)	Quadratic	c_{24}, $b_{23} = 2{\cdot}72$
8	59F8, $a_{28} = 0{\cdot}965$ (300 bands)	Linear	$c_{25} = 1{\cdot}25$
9	60F5, $a_{29} = 1{\cdot}000$ (311 bands)	Quadratic	c_{21}, $b_{24} = 11{\cdot}7$

The numbers of bands (in brackets) refer to the number from the telomere of 2L, in the numbering system of Charlesworth *et al.* (1992*a*).

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

Table 3. *Parameters of chromosome 3*

Region	Proximal boundary (2L) or distal boundary (2R)	Recombination function	Polynomial coefficients
3L			
1	62A12, $a_{31} = 0{\cdot}039$ (15 bands)	Quadratic	$c_{31} = 0{\cdot}022$, $b_{31} = 10{\cdot}6$
2	71A1, $a_{32} = 0{\cdot}277$ (106 bands)	Linear	$c_{32} = 1{\cdot}71$
3	77E1, $a_{33} = 0{\cdot}394$ (151 bands)	Quadratic	$c_{33} = 0{\cdot}038$, $b_{32} = 3{\cdot}33$
4	80A1, $a_{34} = 0{\cdot}441$ (169 bands)	Linear	c_{33}
3L/3R (pericentric region)			
5	81F1–2, $a_{35} = 0{\cdot}444$ (171 bands)	No recombination	
3R			
6	84B1, $a_{36} = 0{\cdot}493$ (189 bands)	Linear	$c_{34} = 0{\cdot}039$
7	89F4, $a_{37} = 0{\cdot}695$ (266 bands)	Quadratic	c_{34}, $b_{33} = 2{\cdot}63$
8	99F1, $a_{38} = 0{\cdot}971$ (372 bands)	Linear	$c_{35} = 1{\cdot}74$
9	100F5, $a_{39} = 1{\cdot}000$ (383 bands)	Quadratic	c_{31}, $b_{34} = 19{\cdot}4$

The numbers of bands (in brackets) refer to the number from the telomere of 3L, in the numbering system of Charlesworth *et al.* (1992*a*).

The euchromatic portion of chromosome 3 includes the polytene divisions on the left arm from 61 to 80C2 (166 bands in our numbering system), and polytene divisions 81–100 on the right arm (214 bands). In addition, there are three bands proximal to 80C2 that are visible in the Lefevre photographic map, which brings the total for 3L to 169 bands, giving a total for the whole chromosome of 383. Physical locations are expressed as proportions of this total, reckoning from the tip of 3L. Chromosome 3 is divided into regions in a similar way to chromosome 2, with boundaries between regions being denoted by $a_{3i} (i = 1–9)$, and the general scheme for relating recombination rate to physical location is the same as for chromosome 2 (Table 3). The justifications for the numerical values of the coefficients are given in the Appendix, Section (iii).

(vi) *Calculating the effect of background selection on a neutral locus*

Given the above assumptions, it is straightforward to calculate the frequency of recombination between a neutral locus at a specified location y and a selected locus at position z, for a given chromosome. This is done by determining the map distance for the coordinate pair (y, z), and using the usual additive rule to combine map distances over all the regions with different recombination parameters that separate them. The map distances concerned are computed using the above formulae for the appropriate chromosomal regions; a mapping function such as eqn (4) is used to obtain the corresponding recombination frequencies. Numerical integration of eqn (2) is then used to determine the ratio of the nucleotide site diversity under background selection to the classical neutral value for the neutral locus in question, due to conventional detrimental mutations. Numerical integration of eqn (3*a*) is used to obtain the corresponding effect of background selection due to TEs. The net effect of background selection is given by the product of these numbers. A picture of the expected effects of background selection as a function of chromosomal location can be obtained by repeating these calculations for a range of y values between 0 and 1. FORTRAN programs to carry out these computations are available on request.

3. Results

(i) *Results of the 'standard' chromosome model and comparisons with the data*

The predictions of the effect of conventional mutations for the three major chromosomes are obtained using the gamma distribution of mutational effects, with the parameter values described in Section 2(i). The background selection effects of TEs are obtained from eqn (3*a*), using the densities of TEs given in Section 2 (ii). The mapping function of eqn (4) is used to describe recombination within chromosome arms. The coefficient of coincidence across the centromere for autosomal regions close to the centromere is assumed to be 10, in rough accordance with published estimates (Green, 1975; Sinclair, 1975; Denell & Keppy, 1979). The other recombination parameters are as in Table 1–3.

Figs. 2–4 shows the predicted values of the nucleotide diversities relative to the classical neutral values (π/π_0), for neutral loci distributed along the X, second and third chromosomes. The π/π_0 values are displayed, taking into account the effects of conventional mutations and TEs separately, as well as their joint effects. Scaled empirical estimates of $\Theta = 4N_e v$ obtained from data on loci surveyed in natural populations of *D. melanogaster* are also displayed for loci distributed over all the major chromosomes, providing a measure of the levels of nucleotide site diversities for the loci concerned (note that N_e is in general different for X-linked and autosomal loci:

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

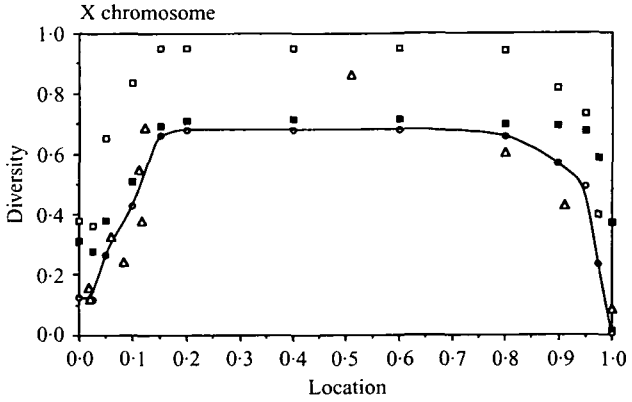

Fig. 2. The observed and expected values of DNA variability as a function of location on the X chromosome of *D. melanogaster*. The expected values are calculated for the 'standard' model of recombination and selection. A gene position is shown as the proportion of the total length of the euchromatin from the telomere to the locus in question. The open and filled squares are the predicted values of π/π_0 under background selection caused by TEs and by conventional mutations, respectively (see Sections 2(i) and 2(ii) of the text for details). The open circles joined by the interpolated curve are the predicted values of π/π_0 under background selection from the combined effect of both forces. The open triangles indicate points obtained from RFLP or SSCP studies of variability within North American or Japanese populations of *D. melanogaster*, scaled as described in the text. Whenever two studies of the same locus were available, the study with the larger number of sites surveyed, or a study of a US population in preference to one on another continent, was used to ensure maximum comparability of the results. The second criterion took precedence over the first. If more than one population from the same continent was studied, the unweighted mean value of the Θ values for each population was used. In order of position, the loci are: *y–ac–sc* (Martín-Campos *et al.* 1992), *su(s)* and *su(w^a)* (Aguadé *et al.* 1994); *Pgd* (Begun & Aquadro 1991), *z*, *tko* (Aguadé *et al.* 1989*b*), *per* (Begun & Aquadro, 1991), *w* (Miyashita & Langley, 1988, as re-analysed by Begun & Aquadro, 1993), *v* and *f* (Miyashita & Langley, 1994), *Zw* (Miyashita, 1990) and *su(f)* (Langley *et al.* 1993).

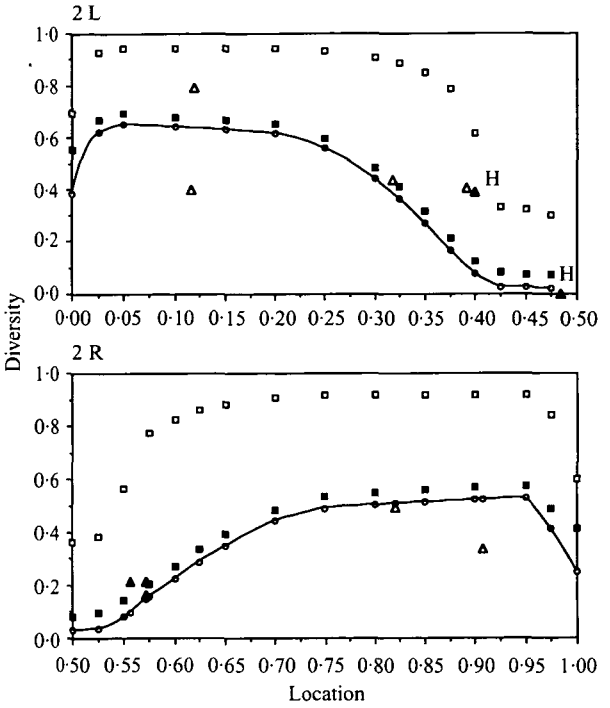

Fig. 3. The observed and expected patterns of DNA variability as a function of location on the second chromosome of *D. melanogaster*. The expected values are calculated for the 'standard' model. A gene position is shown as the proportion of the total length of the euchromatin, from the telomere of 2L to the locus in question. The open and filled triangles indicate the scaled data points obtained from RFLP and sequencing studies, respectively. The symbol H indicates samples formed from pooling strains from around the world, which may give overestimates of sequence diversity within populations. Other symbols are as in Fig. 2. In order of position, the loci are: *Gpdh* (Takano *et al.* 1991), *Mst26A* (Aguadé *et al.* 1992), *Adh* (Takano *et al.* 1991), *Top2* (Palopoli & Wu, unpublished, cited in Kreitman & Wayne, 1994), *Ddc* (Aquadro *et al.* 1992), *cta* (Kreitman & Wayne, 1994), *Gapdh-1* (McDonald & Kreitman, unpublished, cited in Kreitman & Wayne, 1994), *LcpΨ* (Pritchard & Schaeffer, unpublished), *Pgi* (McDonald & Kreitman, unpublished, cited in Kreitman & Wayne, 1994), *Amy* and *Pu* (Takano *et al.* 1991).

Caballero, 1995; Nagylaki, 1995). Estimates of Θ from the numbers of segregating sites in samples were used where available, in preference to estimates from numbers of pairwise differences between sequences, in view of the superior statistical properties of this estimator (Kreitman, 1991). (Background selection is expected to produce only a small bias in this estimate: Charlesworth *et al.* 1995). Scaling was carried out by dividing each Θ estimate for a given chromosome arm by the mean Θ value for that arm, and multiplying by the mean predicted π/π_0 value for the chromosome arm (linear interpolation between directly computed values was used when necessary). This ensures that the mean scaled Θ value for an arm is equal to the mean π/π_0 value, and allows the theoretical and observed relative diversity levels at different chromosomal locations to be compared visually.

Fig. 5 shows plots of the scaled Θ values against the predicted π/π_0 values with both conventional mutations and TEs for all the data, and for the X chromosome alone (which had the highest density of loci studied). If the model is correct, the slope of a fitted straight line should be 1, and the intercept 0. For the pooled data, the least-squares fit of a linear equation is quite good (the squared correlation coefficient, r^2, which gives the proportion of variance in Θ explained by the line, equals 0·68), but the slope is $0{\cdot}72 \pm 0{\cdot}08$ and the intercept is $0{\cdot}10 \pm 0{\cdot}03$, indicating that the data depart significantly from the predictions of the model. For the X chromosome, however, there is a very good fit to the model ($r^2 = 0{\cdot}81$, with an intercept of $0{\cdot}021 \pm 0{\cdot}07$ and a slope of $0{\cdot}96 \pm 0{\cdot}16$).

Several general points can be noted about the results. First, it is clear that even loci in the most highly recombining middle sections of the chromosomes have expected diversities that are substantially

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

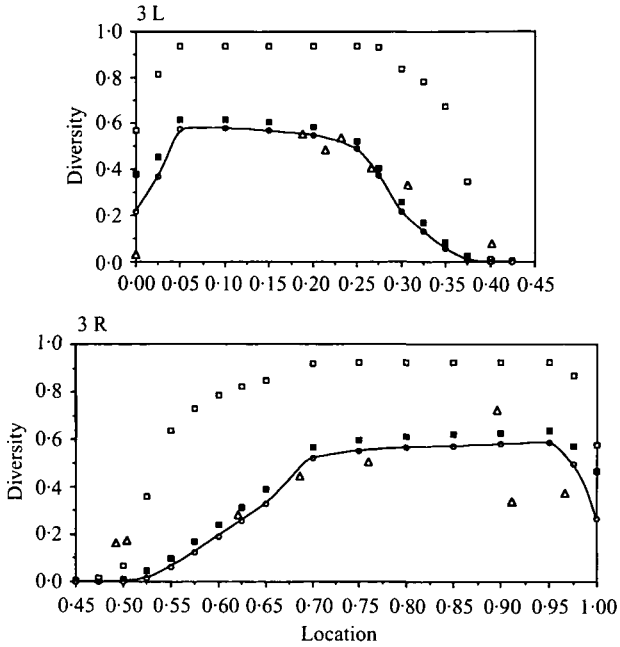

Fig. 4. The observed and expected patterns of DNA variability as a function of location on the third chromosome of *D. melanogaster*. The expected values are calculated for the 'standard' model. A gene position is shown as the proportion of the total length of the euchromatin, from the telomere of 3L to the locus in question. The open triangles indicate scaled data points obtained from an RFLP study of a US population by Kindahl & Aquadro (unpublished). Other symbols are as in Fig. 3. In order of position, the loci shown are: *Lsp1-γ*, *Hsp26*, *Sod*, *Est-6*, *fz*, *tra*, *Pc*, *Antp*, *Gld*, *ry*, *Ubx*, *Rh3*, *E(spl)*, *Tl*, *Mlc2*.

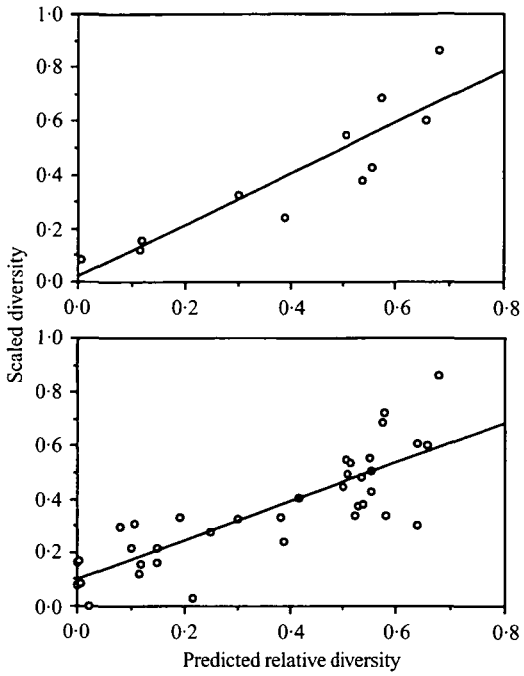

Fig. 5. The plot of the scaled observed Θ values against the corresponding scaled predicted π/π_0 values (see text for method of scaling), with the least-squares fitted straight lines. The lower panel shows the pooled data for all chromosome arms ($y = 0{\cdot}099 + 0{\cdot}722x$; Kendall rank correlation $\tau = 0{\cdot}65$, $P < 0{\cdot}001$). The upper panel is for X chromosomal loci ($y = 0{\cdot}018 \pm 0{\cdot}955x$; Kendall rank correlation $\tau = 0{\cdot}86$, $P < 0{\cdot}001$).

below the classical neutral values, as also found by Hudson & Kaplan (1995). This is mainly due to the effects of conventional mutations rather than TEs. For the X chromosome, the maximum relative diversity is about 68%; for the major autosomes, it lies between about 53% (2R) and 65% (2L). Thus, no locus in *D. melanogaster* is immune to the effects of background selection. Secondly, relatively steep gradients in diversities are expected, especially when the effects of TEs are taken into account. These gradients are not simply artefacts of the discontinuities in the functions relating map distance to physical distance, as the turning points in the curves for π/π_0 do not coincide with the positions of the discontinuities. The gradients are particularly noticeable at the tip and base of the X chromosome; the effects of recombination reduction at the tips of the major autosomes are less marked and the gradients at the bases of the autosomes are shallower than for the X. Nevertheless, extremely low values of genetic diversities for loci close to the centromere are expected, as indeed has been found for the X chromosomal locus *su(f)* (Langley *et al.* 1993) and the 2L locus *cta* (Kreitman & Wayne, 1994).

Thirdly, the peaks in expected variability for the autosomes occur at locations which are at about 10% of the distance from the telomere to the centric heterochromatin for the chromosome arm in question, indicating a surprisingly long-range expected effect of the pericentric restriction of recombination, as also found by Hudson & Kaplan (1995). In all five cases, the observed maximal values of diversity are in regions where expected diversity is close to maximal. Fourthly, there is a tendency for the observed values for the proximal regions of the autosomal arms to exceed the predicted values, while several distal values fall noticeably below the predictions, which presumably causes the discrepancies between the predicted and observed slope and intercept for the pooled data noted in connection with Fig. 5. Possible reasons for these discrepancies will be discussed later. Given the considerable sampling errors in diversity estimates (Kreitman, 1991), some of them may well simply be random in origin.

(ii) *Sensitivity of the predictions to changes in the parameters*

The effects of several kinds of deviations from the standard model were investigated, to test the robustness of the predictions. First, the effect of reducing variation in the strength of selection was investigated, by increasing the parameters of the gamma distribution of selection coefficients to values where the variance is negligible, but keeping the mean the same.

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

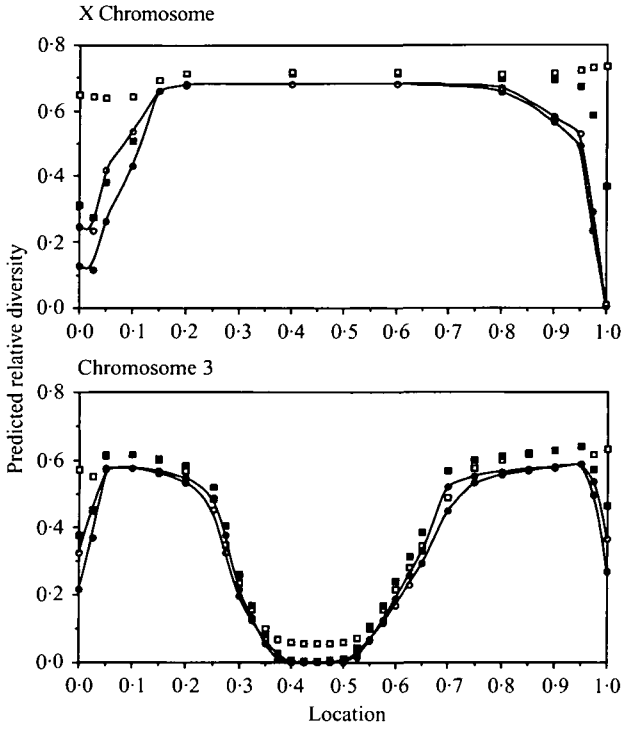

Fig. 6. The effect of removing variation in selection coefficients on the predicted values of π/π_0, for the X and third chromosomes. The open and filled squares are the values for background selection due to conventional mutations alone, without variation in the selection coefficients and for the standard model, respectively. The open and filled circles are the corresponding net effects of conventional mutations and transposable elements.

The results are shown in Fig. 6, which displays both the predictions for the effects of conventional mutations alone and their joint effects with TEs, for the X and third chromosomes. Because there are no weakly selected deleterious mutations with these parameter values, there is less patterning of variability across the chromosomes in response to changes in the local recombinational environment, as expected from the general theoretical results of Hudson & Kaplan (1994, 1995) and Nordborg *et al.* (1996). An increase in diversity at the very ends of the chromosomes, despite the associated reduction in recombination, is more apparent when there is no contribution from weakly selected mutations, especially for the centromere of the X chromosome. The explanation of this edge effect is given by Nordborg *et al.* (1996). In the pericentric regions of the autosomes, the effect of conventional mutations on variation is roughly tenfold less than with the standard parameters (e.g. for chromosome 3, π/π_0 is approximately 0·04 instead of 0·004 for the most proximal loci modelled). Despite this weakening of the telomeric and centromeric reductions in diversity, there are still very marked reductions in variation at the bases of the chromosomes even without TEs, except for the X chromosome.

Secondly, the effect of changing the mapping function was investigated. Haldane's (1919) mapping function, which assumes no interference, provides an obvious alternative to the function used in the standard model. The use of Haldane's function to describe recombination within chromosome arms gave only very slight differences from the previous results when all other parameters were held constant, consistent with the findings of Nordborg *et al.* (1996) on the effects of mapping functions in the case of uniform rates of recombination across the chromosome. For the X chromosome, for example, the maximal diversity under the conventional mutation model with the standard mapping function was 0·71 compared with 0·70 with Haldane's function, and the pattern of change in diversity along the chromosome was similar in both cases. The slightly greater reduction in diversity levels with no interference reflects the fact that recombination distances increase more slowly with map distance with Haldane's function. Results using Kosambi's (1944) used mapping function were also similar to those with the standard one. The details of the mapping function therefore seem to be of minor importance.

Thirdly, the coefficient of coincidence for proximal regions on either side of the centromere was made equal to 1 instead of 10, to examine the effect of removing negative interference across the centromere. A very small increase in diversity close to the centromere resulted, reflecting the effectively higher frequency of recombination per unit map distance. This effect is far too small to be detectable experimentally.

Fourthly, the effect of changing the recombination parameters was examined, for the case of the X chromosome. One rationale for this was that inversion heterozygosity on one chromosome tends to increase recombination frequencies on other chromosomes, particularly in the telomeric and centromeric euchromatin (Lucchesi, 1976). The X chromosome lacks inversion polymorphism, but is exposed to the effects of autosomal inversion heterozygosity. The frequency of X chromosome recombination in nature will therefore be greater than that measured on standardized genetic backgrounds.

Accordingly, the first variant model of the X allows for a longer region where recombination continues to increase at the tip (*Notch* (3C7, 0·128 of the distance from the telomere) is taken as the boundary of region 2 in Table 1 instead of *w*). In accordance with the recombination frequencies measured in the presence of heterozygotes for the common autosomal inversions *In*(2*L*)*t* and *In*(3*R*)*P* by *Sniegowski et al.* (1994), the distance *y*–*N* is taken as 2 cM, giving a revised value of 1·51 for the quadratic coefficient b_{11} relating recombination frequency to physical distance in region 2. This change in the assumptions greatly increases recombination frequencies among the more distal loci in region 2, which is probably over-generous in view of the lack of effects of the common autosomal inversions on recombination in this region (Sniegowski *et al.* 1994). In accordance with the evidence that the

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

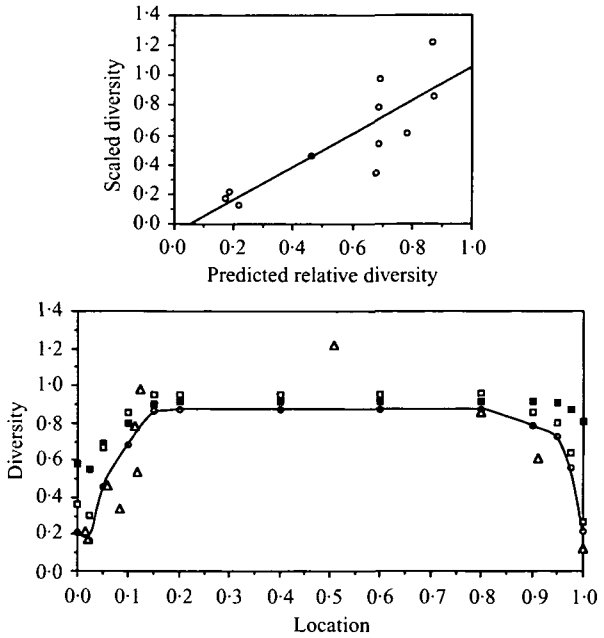

Fig. 7. The lower panel shows the predicted values of π/π_0 and the observed scaled Θ values, for the X chromosome variant model with high rates of distal and proximal recombination (see text for details). The symbols are as in Fig. 2. The upper panel is the plot of the scaled Θ values against π/π_0, with the least-squares fitted straight line ($y = -0{\cdot}059 + 1{\cdot}103x$; Kendall rank correlation $\tau = 0{\cdot}71$, $P < 0{\cdot}01$).

middle section of the X is little affected by autosomal inversion heterozygosity, the only other changes were made in the proximal regions of the X. The assumption that 30% of the population are heterozygous for autosomal inversions is reasonable for populations at intermediate latitudes Lemeunier & Aulard, 1992). In addition, a threefold effect of such heterozygosity on recombination frequency at the base of the X seems to be an upper limit to those observed (Lucchesi, 1976; Sniegowski *et al.* 1994). A net increase in recombination rates of a factor of 1·6 over those assumed earlier thus probably provides an upper bound to what might occur in North American populations. The map distances for regions 4 and 5 are accordingly increased by this factor; in addition, recombination is assumed to extend all the way proximally to the locus of *su(f)*, following the quadratic model of Table 1 with the coefficients adjusted to take into account the expanded map distances. The plot of predicted relative diversity and scaled diversity against location, and the plot of scaled diversity against predicted relative diversity for the X, are shown in Fig. 7. While it is clear that the fit is less good than for the standard model (r^2 is only 0·70 instead of 0·81 for the standard model), there is still relatively good agreement.

The second variant model of the X chromosome is intended to explore the possibility that essentially no recombination occurs between the telomere of the X chromosome and the locus of *y*, as opposed to the low level of recombination assumed here (see Appendix, Section (i)). This seems worthwhile, in view of the very slender evidence on this point (Dubinin *et al.* 1937). If this modification is made, there is only a minor change in the fit of the observed scaled values to the values predicted by the joint effects of TEs and conventional mutations ($r^2 = 0{\cdot}80$). This model predicts, however a much steeper gradient of diversity values near the tip of the X than under the standard model; the ratio of the minimum of π/π_0 (at the telomere) to the maximum value for the chromosome (proximal to *w*) is 0·04 instead of 0·17. This effects is due entirely to a reduction in the distal diversity; π/π_0 proximal to *w* is hardly changed. The gradient in π/π_0 between the telomere and *y* is nearly flat (π/π_0 changes from 0·028 to 0·036, but increases steeply after that).

4. Discussion

The results described above suggest that the background selection model can account for most features of the observed relations between chromosomal location and relative levels of DNA sequence variation in *D. melanogaster*. Hudson & Kaplan (1995) have reached a similar conclusion for loci on chromosome 3, on the basis of a model which does not allow for the effects of TEs. This finding is in contrast to the conclusions of Charlesworth *et al.* (1993), on the basis of a much cruder adjustment for the effects of recombination. If the predictions of these newer models can be taken seriously, it seems that it is unnecessary to invoke factors such as selective sweeps to explain most of the reduced variation associated with regions of reduced recombination.

This raises the question of the robustness of the models to changes in the imperfectly known parameters on which they are based. The relative lack of sensitivity of the predictions to the mapping function used within chromosome arms, and to coincidence across the centromere, is encouraging (Section 3(ii)). But the predictions are very sensitive to changes in the recombination parameters in regions of reduced crossing over, as may be seen from the three versions of the model of the X chromosome compared in Section 3(ii). More detailed information on the relation between recombination rates and physical position in regions of reduced recombination is thus desirable. The other area of uncertainty concerns the per genome rate of mutation to deleterious alleles, and the nature of the distribution of the heterozygous selection coefficients against these alleles. Given the exponential nature of the dependence of the magnitude of the effects of background selection on the mutation rate (eqn 2), a halving of the mutation rate would dramatically reduce the amount of intrachromosomal patterning of variation, and would mean that background selection could explain only a fraction of the reduction in variation in regions of reduced recombination. There is also a sizeable effect of variation in selection coefficients (Fig. 6), due to the major

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

contribution of weakly selected loci to the response to the local recombinational environment (Hudson & Kaplan, 1994, 1995; Nordborg *et al.* 1996). Although we can be fairly confident that the mutation rate per diploid individual for detrimental alleles per genome in *Drosophila* is at least 1, as assumed here, we have little information on the form of the distribution of mutant effects on fitness (Keightley, 1994). Further evidence is badly needed.

There are other uncertainties about the details of the expected patterns of variation which merit comment. Fig. 2 suggests that there is much closer agreement between the observed and predicted patterns of variation on the X chromosome when TEs are included in the predictions than when they are omitted. However, there is little quantitative support for this conclusion; when scaled diversity values are plotted against the predictions based on conventional mutations, r^2 is only slightly lower (0·75) than when TEs are included (0·81). A somewhat similar conclusion holds for chromosome 3 (the heterogeneity of the sources of data for chromosome 2 means that a detailed analysis is probably not meaningful for this chromosome): r^2 is 0·63 for the full model, and 0·53 for the model that includes only conventional mutations. Nevertheless, omission of the effects of TEs means that it is difficult to account for the extremely low variability observed for the proximal X chromosome locus *su(f)* (Langley *et al.* 1993). When TEs are omitted, the scaled diversity value for this locus is 0·11 compared with the predicted value of 0·07, whereas with both factors included the values are 0·09 and 0·005. Similarly, for the tip of the X chromosome the ratio of the predicted values for the two boundary loci *y–ac–sc* and *w* is 0·21 under the standard model including TEs, but is 0·47 if TEs are excluded (and close to 1 if all mutations have the same effect), compared with an observed value of 0·22. This suggests that the presence of weakly selected deleterious entities, such as TEs, is needed to account fully for the gradients or diversity at the tip and base of the X, despite the lack of strong formal statistical support from regression analysis. This is even more evident for the model with relatively high recombination rates at the tip and base of the X (Fig. 7).

This conclusion about the role of TEs in producing extremely low levels of variation in regions of very low recombination is reinforced by the case of chromosome 4. In *D. melanogaster*, the fourth chromosome locus ci^D showed no variation in ten chromosomes sampled from an Illinois population (Berry *et al.* 1991). In view of the small size of the fourth chromosome, this lack of variation is impossible to explain by the background selection model on the basis of conventional mutations alone (Charlesworth *et al.* 1993), and has widely been interpreted as an example of a selective sweep. But data on the distribution of nine TEs on thirteen fourth chromosomes isolated from a natural population in Maryland indicates a mean copy number per chromosome of 3·2 (Charlesworth *et al.* 1992*b*). There is a good fit to a Poisson distribution (the variance is 3·5), so that the frequency of chromosomes free of members of these families can be estimated as $e^{-3.2} = 0{\cdot}04$, where is the value of π/π_0 for these families alone. There is an additional factor of 0·78 from conventional mutations (Charlesworth *et al.* 1993). From the data discussed in Section 2(ii), it also seems likely that the abundance of all element families is about twice that for the set studied on the fourth chromosome, so that the net predicted value of π/π_0 from all sources is $1{\cdot}3 \times 10^{-3}$, which is consistent with the observed lack of variation at ci^D.

But the model may exaggerate the effect of TEs at the autosomal bases, especially for chromosome 3 (Fig. 4). This is because very weakly selected deleterious mutations may have smaller effects on diversity than predicted by eqns (2) and (3), especially when the effective population size is relatively small, since they tend to persist in the population for a long time before elimination (Charlesworth *et al.* 1993; Nordborg *et al.* 1996). Because the effect of background selection can at least in part be viewed as reflecting a reduction in effective population size, the presence of a class of strongly selected mutations may reduce the effectiveness of background selection caused by more weakly selected mutations, since the latter perceive a lower effective population size than when the former are absent (Charlesworth, 1994; Nordborg *et al.* 1996). If all conventional mutations are assumed to have a heterozygous selection coefficient of 0·02, π/π_0 for the base of chromosome 3 is expected to be only about 0·04, if the mutation rate is kept at 1. In consequence, TEs or weakly selected conventional mutations in this region may make much smaller contributions to background selection, leading to an over-prediction of the effects of background selection if their effects are included in the models. It is thus likely that the correct prediction for regions 4–6 of chromosome 3 is intermediate between the predictions with and without the effects of TEs (Fig. 4). This problem is less serious for chromosome 2, where the predicted π/π_0 value at the centromere from strongly selected mutations is 0·14, and it can be ignored for the base of the X and for the tips of the chromosomes (Fig. 6).

There is, however, a difficulty in comparing observed and expected results for the autosomes. This arises from the fact that, as mentioned in Section 3(ii), there is extensive inversion polymorphism for all four autosomal arms in *D. melanogaster* populations (Lemeunier & Aulard, 1992). The presence of such polymorphism has two opposing effects on neutral or nearly neutral variations. First, the extreme reduction of meiotic exchange in inversion heterozygotes, both within the limits of the inverted section and in non-inverted sections of the same chromosomal arm (Roberts, 1976), means that there is a possibility that

chromosome arms carrying a given inversion may gradually diverge genetically from other gene arrangements, giving rise to an elevated level of polymorphism when arrangements are pooled. There is abundant evidence for linkage disequilibrium between allozyme markers and inversions in many *Drosophila* species (Krimbas & Powell, 1992), and some evidence for associations of nucleotide site variants with gene arrangements in *D. melanogaster* in the few studies where relevant data have been obtained (Aquadro *et al.* 1986, 1992; Aguadé, 1988; Wesley & Eanes, 1994). The degree of divergence is likely to be greatest for loci close to inversion breakpoints, since exchange between arrangements by gene conversion in heterokaryotypes is mostly severely reduced in these regions (Chovnick, 1973; Krimbas & Powell, 1992).

But hitch-hiking of neutral variants by a new inversion opposes this increase in variation, since it means that arrangements of recent origin will lack variation, as is often observed (Aquadro *et al.* 1986, 1992; Aguadé, 1988; Krimbas & Powell, 1992). In addition, suppression of crossing over in heterokaryotypes means that selective sweeps and background selection will tend to reduce variation within relatively rare inversions, which occur predominantly as heterozygotes (Wesley & Eanes, 1994). An exception to these phenomena may be provided by loci that are close to the pericentric heterochromatin, for which there is some evidence for increased rates of recombination when distal inversions are heterozygous (Payne, 1924; Sturtevant, 1931; Grell, 1962), particularly in the presence of inversion heterozygosity on other chromosomes (Roberts, 1962).

These factors makes it difficult to make accurate predictions about variation on the autosomes. Since not all the studies of variation used here scored inversions, and in one case rare inversions were deliberately excluded (Takano *et al.* 1991), no attempt has been made to partition variation between different arrangements. Inversions are comparatively rare in the North American and Japanese populations used in these studies (Lemeunier & Aulard, 1992), so that it may well be that the expected effect of inversion polymorphism on autosomal variation is quite small. It is possible, however, that divergence between arrangements could contribute to the generally higher levels of variation on the autosomes compared with X chromosome, noted by Aquadro *et al.* (1994) and Moriyama & Powell (1996). Further theoretical and empirical investigation of this question is needed.

Of course, different selective constraints on the very disparate loci used in these studies may also influence the estimates of variability, complicating the interpretation of apparent discrepancies. In some cases, such as *Mst26A* (Aguadé *et al.* 1992) and *Adh* (Hudson *et al.* 1987), there is clear evidence for selectively maintained polymorphism, which would further distort the picture. One use for comparisons of the observed and expected patterns of variation is to suggest lines of further research to clarify large discrepancies for individual lock. The use of classes of variant which are likely to be neutral or nearly neutral would assist in this. Loci which lie well below the expected curve, such as *T1* and *Mlc2* in the distal section of 3R (Fig. 4), may be candidates for recent selective sweeps in their neighbourhood, for example. Searches for evidence of departure of allele frequency spectra from neutral expectation (Braverman *et al.* 1995; Simonsen *et al.* 1995), or for strong linkage disequilibrium with nearby loci, might shed light on the causes of these discrepancies.

The question remains of how to distinguish the results of background selection from selective sweeps or fluctuating selection in producing the patterns of variation discussed here. These models all predict reduced variation in regions of restricted recombination, and are of course not mutually exclusive. Instances of greatly reduced variation in particular chromosomal regions in some populations but not others seems to provide evidence for selective sweeps due to local adaptation (e.g. *y–ac–sc*: see Martín-Campos *et al.* 1992), and are not expected under background selection. Searches for statistically significant departures of allele frequency spectra from neutral expectation, which are much more likely to be produced by selective sweeps than by background selection or fluctuating selection, may also help to discriminate between alternative explanations of reduced diversity levels in regions of restricted recombination (Braverman *et al.* 1995; Simonsen *et al.* 1995; Charlesworth *et al.* 1995). To date, there are few examples of such departures (Braverman *et al.* 1995; Charlesworth *et al.* 1995), suggesting that selective sweeps have probably not played a major role in causing reduced variability in *D. melanogaster*.

Appendix

(i) *Parameters of the X chromosome*

Region 1: Extreme tip. The *y–ac–sc* complex is located at the proximal end of this region, in 1B1–4. Evidence is scanty on the amount of recombination in this region, but it is close to zero. Padilla & Nash (1977) found no crossovers out of 60000 chromosomes between *cin* (cytological position 1A7) and *y* (1B3), which are separated by two bands on the Bridges' polytene map, representing at least 40 kb of DNA (Heino *et al.* 1994). The data of Dubinin *et al.* (1937) on recombination within the *sc* locus suggest a frequency of about 1×10^{-6} crossovers per kilobase in this region (Aguadé *et al.* 1989*a*). But the relevant crosses were carried out on a background of heterozygosity for *In*(2*L*) (2*R*) *Cy* (Dubinin *et al.* 1937), so that this estimate is probably somewhat inflated, as Redfield (1955) found that recombination at the tip of the X is increased between two and threefold by the presence of this inversion. Half this crossover rate

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

might be reasonable for this region, as the most common naturally occurring second and third chromosome inversions appear to have little effect on telomeric recombination on the X (Sniegowski *et al.* 1994). Given that $a_{11} = 0{\cdot}017$, and that the non-heterochromatic X chromosome is about 22600 kb in size (Heino *et al.* 1994), the total map length of the region from the telomere of the X to *y* is probably at most about 0·019 cM.

In the absence of other information, I assume that crossover rates are independent of map position within this region, so that the map distance between a pair of loci is related to their physical distance *z* by the linear equation

$$l_{11}(z) = c_{11}\, z, \qquad \text{(A 1)}$$

where $c_{11} = 1{\cdot}93 \times 10^{-4}/0{\cdot}017 = 0{\cdot}011$.

Region 2: Proximal part of the tip. The available data suggest that recombination is highly suppressed between *y* and *w*, but that crossover rates per nucleotide proximal to *w* and distal to *f* are roughly independent of location (Redfield, 1955; Lefevre, 1971; Sniegowski *et al.* 1994). The proximal boundary of the proximal part of the tip of the X is therefore taken to be the locus of *w* (3C2). Estimates of the frequency of recombination in this region vary considerably; for example, the conventional map position of 1·5 for *w* is much greater than the estimates of the *y*–*w* distance obtained by Redfield (1955). Similarly, the estimate of Sniegowski *et al.* (1994) for the *y*–*pn* intervals is smaller than the standard map value (*pn* is at 2E2–3). There is evidence that the standard map of the X was based on crosses in which heterozygous inversions were present on chromosome 3 (Ashburner, 1989, p. 453), which may explain these discrepancies. For the sake of conservatism, I shall use the standard map value, giving an estimate of 1·48 cM for *y*–*w*.

It is clear that the crossover frequency per base pair increases proximally within this region (Lefevre, 1971; Sniegowski *et al.* 1994); the simplest representation of this is to assume that the density of crossover events increases linearly with distance from the distal boundary of this region. For consistency with region 1, the density at the distal boundary is set equal to c_{11}. The map distance between two loci, at positions z_1 and z_2 ($z_2 > z_1$) within the region, is obtained from the integral of the density function, and hence is a quadratic function

$$l(z_1, z_2) = c_{11}(z_2 - z_1) + b_{11}\{(z_2 - a_{11})^2 - (z_1 - a_{11})^2\}, \qquad \text{(A 2)}$$

where the value of the constant b_{11} is obtained by equating the map distance between *y* and *w* to the prediction of this formula for $z_1 = a_{11}$ and $z_2 = a_{12}$ ($b_{11} = 1{\cdot}21$). Aguadé *et al.* (1994) quote values of 10^{-4} and $3{\cdot}5 \times 10^{-3}$ for recombination frequencies between *y* and the more proximal loci *su*(*s*) and *su*(w^a), respectively. These agree quite well with the values predicted by eqn (A 2) ($8{\cdot}2 \times 10^{-5}$ and $2{\cdot}6 \times 10^{-3}$), giving some confidence in this approach.

Region 3: Middle region. The proximal boundary of this region is taken to be the location of *f* (15F1–3), since this is where recombination starts to decrease again (Lefevre, 1971). Map distance within this region is assumed to be related to physical separation by a linear function with constant of proportionality c_{12}. Using the standard map distance between *w* and *f* (55·2 cM), we have $c_{12} = 0{\cdot}817$). This is much larger than the slope of eqn (A 2) at a_{12} (0·268), suggesting that the rate of exchange continues to increase proximal to *w*. The relatively large standard map distance (1·5 cM) between *w* and *N*, which are only one Lefevre map band apart, suggests that there is an abrupt local increase in recombination rate just proximal to *w*. Ignoring this increase means that the slope of the relation between genetic diversity and location is underestimated in the distal part of the X.

Region 4: Distal part of proximal region. The rate of crossing over per nucleotide between the locations of *f* and *mal* is about 60% of that in the mid-section (Schalet & Lefevre, 1976). *mal* is located at 19D2–3, which is used as the proximal boundary for this region. There is no firm evidence that crossover frequency per nucleotide decreases proximally here, so that I assume that map and physical distance are proportional, with constant of proportionality c_{13}. For consistency with the recombination function used for region 5 (see below), I take c_{13} to be equal to the slope of the function relating map distance to physical position in region 5, evaluated at $z = a_{14}$. This is consistent with the estimated frequency of crossing over for this region (Schalet & Lefevre, 1976).

Region 5: Central part of proximal region. There is probably a gradient of crossover density from the locus of *mal* towards a near-zero frequency in the β-heterochromatin (Schalet & Lefevre, 1976). Arbitrarily, I assume that the proximal boundary of the region of non-zero crossover frequencies is at band 177, three bands distal to the location of *su*(*f*) at 20F, and well proximal to the end of the euchromatin at 20A2 (band 172). I assume a quadratic relation between map and physical distance, with no linear term and with quadratic coefficient b_{12}. The value of b_{12} is obtained by calibration against the 1·6 cM distance between *mal* and *su*(*f*) (Schalet & Lefevre, 1976).

Region 6: Proximal part of proximal region. Zero recombination is assumed in the remaining 1·1% of the X chromosome distal to *su*(*f*). There is a distance of 0·068 cM between a lethal in 20A and *su*(*f*) (Schalet, 1972), which are separated by six or seven bands in our system of numbering. This is consistent

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

with a very low frequency of crossing over in this region.

(ii) *Parameters of chromosome 2*

Region 1. Tip of 2L. Measurements of recombination frequencies suggest that the map length of this region as defined in Table 2 is at most about 2 cM (Lewis, 1945; Roberts & Evans-Roberts, 1979). There is also evidence for an increase in recombination frequency just proximal to this region (Lewis, 1945; Roberts *et al.* 1985), suggesting that there is a gradient of increasing recombination away from the telomere, but the detailed nature of this is unknown. I use the quadratic function

$$l(z) = c_{21} z + b_{21} z^2. \quad \text{(A 3)}$$

The value of the linear coefficient c_{21} is obtained by assuming that the rate of recombination per kilobase at the tips of the autosomes is the same as that for the tip of the X chromosome. The total recombination frequency for the tip of the X is $c_{11} a_{11}$, and the number of kilobases at the tip of the X is the product of a_{11} and the size of the X in kilobases. The distal rate of exchange per kilobase for the X is thus c_{11} divided by the size of the X. If this is the same as the rate for chromosome 2, c_{21} is equal to the product of c_{11} and the ratio of the size of chromosome 2 to that of the X chromosome, which is about 1·80 (Heino *et al.* 1994), i.e. $c_{21} = 0{\cdot}020$. The value of b_{21} is obtained by equating the total map length of the region to the right-hand side of eqn (A 3).

Region 2: Middle of 2L. The data of Ising (Ashburner, 1989, chap. 11) suggest that map distance is approximately linearly related to physical position in this region with coefficient c_{22}, obtainable from the total map length of the region (38 cM).

Region 3: Distal part of proximal region of 2L. Ising's data suggest a curvilinear relation between map length and physical location in this region, so that I use a quadratic function analogous to eqn (A 2), with linear and quadratic coefficients c_{23} and b_{22}. c_{23} is assumed to be the same as the linear coefficient for region 4 (0·144); the total map length of the region is 14 cM (Ashburner, 1989, p. 454), giving $b_{22} = 4{\cdot}70$.

Region 4: Central part of proximal region of 2L. The distance between *Bl* and *ap*, located at 38A6-E9 and 41B–C respectively, is 1 cM (Sturtevant, 1949), which provides an overestimate for the map length of region 4.

Region 5: Pericentric regions of 2L and 2R. Two euchromatic bands are included in this region, one on each side of the centromere. I assume that no recombination takes place in this region.

Region 6: Central part of proximal region of 2R. This is treated like region 4, except that the map length of this region is approximately 0·7 cM (Ives, 1947).

Region 7: Distal part of proximal region 2R. This is treated like region 3, but with total map length 13·5 cM (Ashburner, 1989, p. 455).

Region 8: Middle of 2R. This is similar to region 2, except that the total map length is 28·5 cM (Ashburner, 1989, p. 455).

Region 9: Tip of 2R. Mapping studies indicate that there is a steep gradient of recombination rate between 59D–E and 60B–C (Ives, 1967; Sato, 1984; Schüpbach & Wieschaus, 1989), with a map length of approximately 1·5 cM for the whole tip region. I use a quadratic function similar to that for the tip of 2L.

(iii) *Parameters of chromosome 3*

Region 1: Tip of 3L. There is gradient of decreasing frequency of recombination towards the tip: the map distance from the *Lsp*-1-γ locus at 61A1 to *fap* at 61F is 0·91 cM (11 bands in my system of numbering), whereas the distance from *fap* to *ve* (62A) is 0·63 cM (3 bands) (Roberts & Evans-Roberts, 1979). The locus of *ve* is 1·7 cM from the tip. The ratio of the sizes of the third and X chromosomes is approximately 1·98 (Heino *et al.* 1994), which enables the value of c_{31} to be obtained in a similar way to c_{21}.

Region 2: Middle of 3L. The location of *ve* and Ising's data (Ashburner 1989, p. 457) indicate that the total map length of this region is 42·5–1·7 = 40·8 cM.

Region 3: Distal part of proximal section of 3L. The map distance over this region is 5 cM, according to Ising's data.

Region 4: Central proximal region of 3L. The data of Green (1975) and Sinclair (1975) suggest that there are about 1×10^{-4} crossovers per band for euchromatin in the proximal third chromosome.

Region 5: Proximal parts of proximal regions of 3L and 3R. The last band of 3L and first band of 3R are assumed to lack crossing over, and define this region. The centromere is assumed to be midway between them.

Region 6: Central part of proximal region of 3R. The estimate of the rate of exchange per band for region 4 is used to obtain the c coefficient for this region.

Region 7: Distal part of proximal region of 3R. The total map length for this region is 11·5 cM (Ashburner, 1989, p. 457).

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

Region 8: Central region of 3R. The total map length is 48 cM (Ashburner, 1989, p. 457).

Region 9: Distal region of 3R. Published data on recombination at the tip of 3R give conflicting values for the same intervals – e.g. Karess & Glover (1989) found the distance between *ca-rod* to be 4·8 cM, and Sturtevant (1956) obtained a value of 2·0 cM for *ca-avd*^{K}. *ca* is at 99B11 while the other two loci are in 100C–D. In the absence of direct evidence on the relation between map and physical distance in this region, I assume that the total map length of salivary division 100 is the same as that for the tip of 3L (1·7 cM), and that there is a quadratic relation between map and physical distance with the same linear coefficient as for 3L.

(iv) *Effect of negative interference across the centromere*

It is assumed that negative interference across the centromere occurs between regions 3 or 4 on the left arm of an autosome, and regions 6 and 7 on the right arm of the same autosome (Sections (ii) and (iii) above). It will also be assumed that loci distal to specified boundary points located in regions 3 and 7 are immune to interference across the centromere. The boundary points were standardly assumed to be located 0·05 of the chromosome distal to the proximal borders of regions 3 and 7. Consider first a neutral site on a given arm, located distally to the boundary point for that arm, and a selected locus that is proximal to the boundary point on the other arm. Let the segment of the chromosome from the neutral site to the boundary point be a, the segment which extends proximally from this to the centromere be b, and the segment from the centromere to the selected locus be c. Let $R(i)$ be the probability of a recombination event in segment i and $R(\bar{\imath})$ be the probability of no recombination in that segment. Then the frequency of recombination between the neutral and selected loci is (neglecting triple crossovers)

$$R(a+b+c) \approx R(a \cap \bar{b} \cap \bar{c}) + R(\bar{a} \cap b \cap \bar{c}) + R(\bar{a} \cap \bar{b} \cap c). \quad \text{(A 4)}$$

Since events in segments a and c are independent, we have

$$R(a \cap \bar{b} \cap \bar{c}) = R(\bar{c}\,|\,\bar{b})\, R(a \cap \bar{b}),$$

where

$$R(a \cap \bar{b}) = R(a) - R(a \cap b)$$

and

$$R(a+b) = R(a) + R(b) - 2R(a \cap b),$$

giving

$$R(a \cap \bar{b}) = \frac{1}{2}\{R(a) + R(a+b) - R(b)\}. \quad \text{(A 5)}$$

We also have

$$R(c) = R(c\,|\,\bar{b})\, R(\bar{b}) + R(c\,|\,b)\, R(b)$$
$$= R(c\,|\,\bar{b})\{1 - R(b)\} + C\, R(b)\, R(c),$$

where C is the coefficient of coincidence for recombination across the centromere.

This yields the relations

$$R(c\,|\,\bar{b}) = R(c)\{1 - C\, R(b)\}/\{1 - R(b)\} \quad \text{(A 6)}$$

and

$$R(\bar{c}\,|\,\bar{b}) = \{1 - R(b)[1 - C\, R(c)] - R(c)\}/\{1 - R(b)\}. \quad \text{(A 7)}$$

Hence,

$$R(a \cap \bar{b} \cap \bar{c}) = \frac{\{R(a) + R(a+b) - R(b)\}\{1 - R(b)[1 - C\, R(c)] - R(c)\}}{2\{1 - R(b)\}}. \quad \text{(A 8)}$$

Similarly,

$$R(\bar{a} \cap b \cap \bar{c}) = R(\bar{c}\,|\,b)\, R(\bar{a} \cap b)$$
$$= \tfrac{1}{2}\{1 - C\, R(c)\}\{R(b) + R(a+b) - R(a)\} \quad \text{(A 9)}$$

and

$$R(\bar{a} \cap \bar{b} \cap c) = \frac{\{1 - \tfrac{1}{2}[R(a) + R(b) + R(a+b)]\} R(c)\{1 - C\, R(b)\}}{\{1 - R(b)\}}. \quad \text{(A 10)}$$

Since the recombination fractions in eqns (A 8–A 10) are all determined by the map distances for the intervals in question, substitution of these relations into eqn (A 4) provides a means of computing the desired recombination frequency for segments a, b and c.

I next consider the case of a neutral site in the same region as before, but with the selected locus located distal to the boundary on the other arm. Segment c is now the region that extends from the centromere to this boundary, and segment d denotes the region between the boundary and the selected locus. Hence

$$R(a+b+c+d) \approx R(a \cap \bar{b} \cap \bar{c} \cap \bar{d}) + R(\bar{a} \cap b \cap \bar{c} \cap \bar{d}) + R(\bar{a} \cap \bar{b} \cap c \cap \bar{d}) + R(\bar{a} \cap \bar{b} \cap \bar{c} \cap d). \quad \text{(A 11)}$$

We have

$$R(a \cap \bar{b} \cap \bar{c} \cap \bar{d}) = R(\bar{d}\,|\,\bar{c})\, R(\bar{c}\,|\,\bar{b})\, R(a \cap \bar{b}). \quad \text{(A 12)}$$

An argument similar to those used above gives

$$R(\bar{d}\,|\,\bar{c}) = \frac{\{1 - \tfrac{1}{2}[R(c) + R(d) + R(c+d)]\}}{\{1 - R(c)\}}. \quad \text{(A 13)}$$

The other components of eqn (A 12) are obtained from eqns (A 5) and A 7).

https://doi.org/10.1017/S0016672300034029 Published online by Cambridge University Press

Similarly,

$$R(\bar{a} \cap b \cap \bar{c} \cap \bar{d}) = \frac{\{1 - CR(c)\}\{R(b) + R(a+b) - R(a)\}\{1 - \frac{1}{2}[R(c) + R(d) + R(c+d)]\}}{2\{1 - R(c)\}} \quad \text{(A 14)}$$

and

$$R(\bar{a} \cap \bar{b} \cap \bar{c} \cap d) = \frac{\{1 - \frac{1}{2}[R(a) + R(b) + R(a+b)]\}\{1 - R(b)[1 - C\,R(c)] - R(c)\}\{R(d) + R(c+d) - R(c)\}}{2\{1 - R(b)\}\{1 - R(c)\}}. \quad \text{(A 15)}$$

For a neutral site that is proximal to the boundary for interference across the centromere on the same chromosome arm, we can proceed as follows, defining segment b as the region between the neutral site and the centromere. If the selected locus is proximal to the boundary on the other arm, the standard mapping formula gives

$$R(b+c) = R(b) + R(c) = 2C\,R(b)\,R(c). \quad \text{(A 16)}$$

If the selected locus is distal to this boundary, then

$$R(b+c+d) \approx R(b \cap \bar{c} \cap \bar{d}) + R(\bar{b} \cap c \cap \bar{d}) + R(\bar{b} \cap \bar{c} \cap d). \quad \text{(A 17)}$$

This equation is analogous to eqn (A 4), and can be evaluated in the same way, if d is substituted for a, and c and b are interchanged.

I thank Charles Aquadro, Jody Hey, Jeffrey Powell, Stephen Schaeffer and Gail Simmons for providing me with their results prior to publication. Nick Barton, Deborah Charlesworth, James Crow, David Begun, Marty Kreitman, Charles Langley, Cathy Laurie, Magnus Nordborg, Janice Spofford, Wolfgang Stephan and an anonymous reviewer made helpful suggestions for improving the manuscript. I am particularly grateful to Richard Hudson for sharing his theoretical results, and for drawing my attention to the role of TEs in background selection. This work was supported by the Darwin Trust of Edinburgh, and by grant DEB93117683 from the National Science Foundation.

References

Aguadé, M. (1988). Restriction map variation of the *Adh* locus of *Drosophila melanogaster* in inverted and non-inverted chromosomes. *Genetics* **119**, 135–140.

Aguadé, M., Miyashita, N. & Langley, C. H. (1989*a*). Reduced variation in the *yellow–achaete–scute* region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615.

Aguadé, M., Miyashita, N. & Langley, C. H. (1989*b*). Restriction-map variation at the *zeste–tko* region in natural populations of *Drosophila melanogaster*. *Molecular Biology and Evolution* **6**, 123–130.

Aguadé, M., Miyashita, N. & Langley, C. H. (1992). Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila melanogaster*. *Genetics* **132**, 755–770.

Aguadé, M. & Langley, C. H. (1994). Polymorphism and divergence in regions of low recombination in *Drosophila*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 67–76. London: Chapman and Hall.

Aguadé, M., Meyers, W., Long, A. D. & Langley, C. H. (1994). Reduced DNA sequence polymorphism in the *su(s)* and *s(w^a)* regions of *Drosophila melanogaster* as revealed by SSCP and stratified DNA sequencing. *Proceedings of the National Academy of Sciences of the USA* **91**, 4658–4662.

Aquadro, C. F., Deese, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1986). Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**, 1165–1190.

Aquadro, C. F., Jennings, R. M. J., Bland, M. M., Laurie, C. C. & Langley, C. H. (1992). Patterns of naturally occurring restriction map variation and linkage disequilibrium in the *Ddc* region of *Drosophila melanogaster*. *Genetics* **132**, 443–452.

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 46–56. London: Chapman and Hall.

Ashburner, M. (1989). *Drosophila. A Laboratory Handbook*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Barton, N. H. (1995). Linkage and the limits to natural selection. *Genetics* **140**, 821–884.

Begun, D. J. & Aquadro, C. F. (1991). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow–achaete* region. *Genetics* **129**, 1147–1158.

Begun, D. J. & Aquadro, C. F. (1992). Levels of natural occurring DNA polymorphism correlate with recombination rate in *Drosophila melanogaster*. *Nature* **356**, 519–520.

Begun, D. J. & Aquadro, C. F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Biémont, C. (1992). Population genetics of transposable DNA elements: a *Drosophila* point of view. *Genetica* **86**, 67–84.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140**, 783–796.

Brooks, L. D. (1988). The evolution of recombination rates. In *The Evolution of Sex* (ed. R. E. Michod & B. R. Levin), pp. 87–105. Sunderland, MA: Sinauer.

Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**, 1007–1011.

Charlesworth, B. (1991). Transposable elements in natural populations, with a mixture of selected and neutral insertion sites. *Genetical Research* **57**, 127–134.

Charlesworth, B. (1994). The effect of background selection against deleterious alleles on weakly selected, linked variants. *Genetical Research* **63**, 213–228.

Charlesworth, B. & Langley, C. H. (1991). Population genetics of transposable elements in *Drosophila*. In *Evolution at the Molecular Level* (ed. R. K. Selander, A. G. Clark & T. S. Whittam), pp. 150–176. Sunderland, MA: Sinauer.

Charlesworth, B. & Lapid, A. (1989). A study of ten transposable elements on X chromosomes from a population of *Drosophila melanogaster*. *Genetical Research* **54**, 113–125.

Charlesworth, B., Lapid, A. & Canada, D. (1992*a*). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* **60**, 103–114.

Charlesworth, B., Lapid, A. & Canada, D. (1992*b*). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. *Genetical Research* **60**, 115–130.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Charlesworth, D., Charlesworth, B. & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632.

Chovnick, A. (1973). Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. *Genetics* **74**, 123–131.

Cobbs, G. (1978). Renewal approach to the theory of genetic linkage: case of no chromatid interference. *Genetics* **89**, 563–581.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila*. Vol. 3c (ed. M. Ashburner, H. L. Carson & J. N. Thomson), pp. 1–35. London: Academic Press.

Denell, R. E. & Keppy, D. O. (1979). The nature of genetic recombination near the third chromosome centromere of *Drosophila melanogaster*. *Genetics* **93**, 117–130.

Dubinin, N. P., Sokolov, N. N. & Tiniakov, G. G. (1937). Crossing over between the genes 'yellow' and 'scute'. *Drosophila Information Service* **8**, 76.

Gillespie, J. H. (1994). Alternatives to the neutral theory. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 1–17. London: Chapman and Hall.

Green, M. M. (1975). Conversion as a possible mechanism of high coincidence values in the centromeric region of *Drosophila*. *Molecular and General Genetics* **44**, 1243–1256.

Grell, R. F. (1962). A new model for secondary nondisjunction: the role of distributive pairing. *Genetics* **47**, 1737–1754.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distance between loci of linked factors. *Journal of Genetics* **8**, 299–309.

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838–844.

Heino, T. I., Saura, A. O. & Sorsa, V. (1994). Maps of the salivary gland chromosomes of *Drosophila melanogaster* *Drosophila Information Service* **73**, 621–738.

Houle, D., Hughes, K. A., Hoffmaster, D. K., Ihara, J. T., Assimacopoulos, S. & Charlesworth, B. (1994). The effect of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. *Genetics* **138**, 773–785.

Hudson, R. R. (1994). How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proceedings of the National Academy of Sciences of the USA* **91**, 6815–6818.

Hudson, R. R. & Kaplan, N. L. (1994). Gene trees with background selection. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 140–153. London: Chapman and Hall.

Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hughes, K. A. (1995). The inbreeding decline and average dominance of genes affecting male life-history characters in *Drosophila melanogaster*. *Genetical Research* **65**, 41–45.

Ives, P. T. (1947). Report of P. T. Ives. *Drosophila Information Service* **21**, 68–69.

Ives, P. T. (1967). Relocation of the *or* locus closer to *pd*. *Drosophila Information Service* **42**, 76.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitch-hiking' effect revisited. *Genetics* **123**, 887–899.

Karess, R. E. & Glover, D. M. (1989). *Rough deal*: a gene required for proper mitotic segregation in *Drosophila*. *Journal of Cell Biology* **109**, 2951–2961.

Keightley, P. D. (1994). The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1–8.

Kosambi, D. D. (1944). The estimation of map distance from recombination values. *Annals of Eugenics* **12**, 172–175.

Kreitman, M. (1991). Detecting selection at the level of DNA. In *Evolution at the Molecular Level* (ed. R. K. Selander, A. G. Clark & T. S. Whittam), pp. 202–221. Sunderland, MA: Sinauer.

Kreitman, M. & Wayne, M. L. (1994). Organization of genetic variation at the molecular level: lessons from *Drosophila*. In *Molecular Ecology and Evolution: Approaches and Applications* (ed. B. Schierwater, B. Streit, G. P. Wagner & R. DeSalle), pp. 157–184. Basel: Birkhäuser.

Krimbas, C. B. & Powell, J. R. (1992). Introduction. In *Inversion Polymorphism in Drosophila* (ed. C. B. Krimbas & J. R. Powell), pp. 1–52. Boca Raton, FL: CRC Press.

Langley, C. H., Montgomery, E. A., Hudson, R. R., Kaplan, N. L. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **52**, 223–235.

Langley, C. H., MacDonald, J., Miyashita, N. & Aguadé, M. (1993). Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the USA* **90**, 1800–1803.

Lefevre, G. (1971). Salivary chromosome bands and the frequency of crossing over in *Drosophila melanogaster*. *Genetics* **67**, 497–513.

Lefevre, G. (1976). A photographic representation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In *The Genetics and Biology of Drosophila* (ed. M. Ashburner & E. Novitski), pp. 31–36. Orlando, FL: Academic Press.

Lemeunier, F. & Aulard, S. (1992). Inversion polymorphism in *Drosophila melanogaster*. In *Drosophila Inversion Polymorphism* (ed. C. B. Krimbas & J. R. Powell), pp. 339–406. Boca Raton, FL: CRC Press.

Lewis, E. B. (1945). The relation of repeats to position effects in *Drosophila melanogaster*. *Genetics* **30**, 137–166.

Lindsley, D. L. & Sandler, L. (1977). The genetic analysis of meiosis in female *Drosophila*. *Philosophical Transactions of the Royal Society of London, Series B* **277**, 295–312.

Lindsley, D. L. & Zimm, G. G. (1992). *The Genome of Drosophila melanogaster*. San Diego, CA: Academic Press.

Lucchesi, J. C. (1976). Inter-chromosomal effects. In *The Genetics and Biology of Drosophila*, vol. 1a (ed. M. Ashburner & E. Novitski), pp. 315–330. New York: Academic Press.

Martín-Campos, J. M., Coméron, J. M., Miyashita, N. & Aguadé, M. (1992). Intraspecific and interspecific variation at the *y–ac–sc* region of *Drosophila simulans* and *Drosophila melanogaster*. *Genetics* **130**, 805–816.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

Miklos, G. L. G. & Cotsell, J. N. (1990). Chromosome structure at interfaces between major chromatin types: *alpha*- and *beta*-heterochromatin. *Bioessays* **12**, 1–6.

Miyashita, N. T. (1990). Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. *Genetics* **125**, 407–419.

Miyashita, N. T. & Langley, C. H. (1988). Molecular and phenotypic evolution of the *white* locus in *Drosophila melanogaster*. *Genetics* **120**, 199–212.

Miyashita, N. T. & Langley, C. H. (1994). Restriction map polymorphism in the *forked* and *vermilion* regions of *Drosophila melanogaster*. *Japanese Journal of Genetics* **69**, 297–305.

Montgomery, E. A., Charlesworth, B. & Langley, C. H. (1987). A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genetical Research* **49**, 31–41.

Montgomery, E. A., Huang, S.-M., Langley, C. H. & Judd, B. H. (1991). Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* **129**, 1085–1098.

Moriyama, E. N. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

Nagylaki, T. (1995). The inbreeding effective population number in dioecious populations. *Genetics* **139**, 473–485.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetical Research* **67**, 159–174.

Nuzhdin, S. V. & Mackay, T. F. C. (1995). The genomic rate of transposable element movement in *Drosophila melanogaster*. *Molecular Biology and Evolution* **12**, 180–181.

Ohnishi, O. (1977). Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effects of polygenic mutations. *Genetics* **87**, 529–545.

Ohnishi, S. & Voelker, R. A. (1979). Comparative studies of allozyme loci in *Drosophila simulans* and *D. melanogaster*. II. Gene arrangement on the third chromosome. *Japanese Journal of Genetics* **54**, 203–209.

Owen, A. R. G. (1951). An extension of Kosambi's formula. *Nature* **168**, 208–209.

Padilla, H. M. & Nash, W. G. (1977). A further characterization of the *cinnamon* gene in *Drosophila melanogaster*. *Molecular and General Genetics* **155**, 171–177.

Payne, F. (1924). Crossover modifiers in the third chromosome of *Drosophila melanogaster*. *Genetics* **9**, 327–342.

Redfield, H. (1955). Recombination increase due to heterologous inversions and the relation to cytological length. *Proceedings of the National Academy of Sciences of the USA* **41**, 1084–1091.

Roberts, D. B., Brock, H. W., Rudden, N. C. & Evans-Roberts, S. (1985). A genetic and cytogenetic analysis of the region surrounding the LSP-1-β gene in *Drosophila melanogaster*. *Genetics* **109**, 145–156.

Roberts, D. R. & Evans-Roberts, S. (1979). The genetic and cytogenetic localization of the three structural genes coding for the major protein of *Drosophila* larval serum. *Genetics* **93**, 663–679.

Roberts, P. A. (1962). Interchromosomal effects and the relation between crossing-over and nondisjunction. *Genetics* **47**, 1691–1710.

Roberts, P. A. (1976). The genetics of chromosome aberration. In *The Genetics and Biology of Drosophila*, vol. 1a (ed. M. Ashburner & E. Novitski), pp. 68–184. New York: Academic Press.

Sato, T. (1984). A new homeotic mutation affecting antennae and legs. *Drosophila Information Service* **60**, 180–182.

Schalet, A. (1972). Crossing over in the major heterochromatic region of the X chromosome in normal and inverted sequences. *Drosophila Information Service* **48**, 111–113.

Schalet, A. & Lefevre, G. (1976). The proximal region of the X chromosome. In *The Genetics and Biology of Drosophila*, vol. 1b (ed. M. Ashburner & E. Novitski), pp. 847–902. London: Academic Press.

Schüpbach, T. & Wieschaus, E. (1989). Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119–1136.

Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Sinclair, D. A. (1975). Crossing over between closely linked markers spanning the centromere of chromosome 3 in *Drosophila melanogaster*. *Genetical Research* **26**, 173–186.

Sniegowski, P. D., Pringle, A. & Hughes, K. A. (1994). Effect of autosomal inversions on meiotic exchange in distal and proximal regions of the X chromosome in a natural population of *Drosophila melanogaster*. *Genetical Research* **63**, 57–62.

Stephan, W. (1995). An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. *Molecular Biology and Evolution* **12**, 959–962.

Sturtevant, A. H. (1931). Known and probably inverted sections of the autosomes of *Drosophila melanogaster*. *Carnegie Institution of Washington Publications* **421**, 1–27.

Sturtevant, A. H. (1949). Sequence of loci near the centromere of chromosome 2. *Drosophila Information Service* **23**, 98.

Sturtevant, A. H. (1956). A highly specific complementary lethal system in *Drosophila melanogaster*. *Genetics* **41**, 118–123.

Takano, T. S., Kusakabe, S. & Mukai, T. (1991). The genetic structure of natural populations of *Drosophila melanogaster*. XXII. Comparative studies of DNA polymorphism in northern and southern populations. *Genetics* **129**, 753–761.

Wesley, C. S. & Eanes, W. F. (1994). Isolation and analysis of the breakpoint sequences of chromosome inversion *In (3L)Payne* in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **91**, 3132–3136.

Wiehe, T. H. E. & Stephan, W. (1993). Analysis of a genetic hitchhiking model and its application to DNA polymorphism data. *Molecular Biology and Evolution* **10**, 842–854.