Heredity **62** (1989) 97–106

Received 17 April 1988

Genetic analysis of X-linked sterility in hybrids between three sibling species of *Drosophila*

Jerry A. Coyne and
Brian Charlesworth

Department of Ecology and Evolution,
The University of Chicago, 1103 East 57th Street,
Chicago, Illinois 60637 U.S.A.

Three morphological markers (*yellow*, *miniature*, and *forked*) are used to to map the location of X-chromosome segments causing male sterility in *Drosophila simulans*/*D. mauritiana* and *D. simulans*/*D. sechellia* hybrids. In both hybridizations at least three sections of the chromosome contain genes with substantial effects on sterility. This represents the maximum genetic divergence detectable with the three markers, suggesting that the X chromosome contains many loci affecting postzygotic reproductive isolation. The tight linkage between some markers and "sterility loci" may be useful in localizing and later cloning genes important in speciation.

INTRODUCTION

The preferential sterility or invariability of the heterogametic sex in species hybrids is a ubiquitous pattern of speciation known as "Haldane's rule" (Haldane, 1922). Genetic analysis of this pattern has produced a second generalization: the genes having the greatest effect on hybrid sterility and inviability are found on the X chromosome. This result is reported by virtually every genetic study of postzygotic reproductive isolation (see Coyne and Orr, 1988 for documentation and Charlesworth *et al.*, 1987 for a theoretical explanation).

The enumeration and mapping of these "sterility" genes is important for several reasons. Some recent theories assert that one or two genes of large effect could initiate speciation (Gould, 1980; Templeton, 1981). This prediction could be tested by examining the number of loci causing postzygotic isolation, particularly in hybrids between recently-diverged species. Moreover, fine-structure mapping of such loci may eventually allow cloning, sequencing, and determination of their normal function. We would like to know, for example, if "sterility genes" produce components of sperm, control the timing of meiosis, or perform some other developmental function having little to do with reproduction.

Several studies have counted or mapped such loci by observing the linkage between hybrid sterility and genetic or chromosomal markers. In backcrosses between *Drosophila buzzatii* and *D. serido*, Naveira and Fontdevila (1986) found that every X-linked segment from *D. serido*, no matter how small, caused male sterility when introgressed into *D. buzzatii*. This implies that hybrid sterility results from many X-linked genes of large effect. At least four X-linked genes are responsible for male sterility in hybrids between the sibling species *D. pseudoobscura* and *D. persimilis* (Lancefield, 1929; Wu and Beckenbach, 1983). In hybrids between *D. simulans* and its two island relatives *D. mauritiana* and *D. sechellia*, each end of the X chromosome has a substantial effect on male sterility, so that at least two loci are involved (Coyne and Kreitman, 1986).

All of these studies revealed the maximum number of "sterility genes" detectable with the genetic techniques used, implying that there may be a large number of *undetected* loci with effects on sterility. Finer-scale genetic analysis is required to map more accurately the known sterility genes. Observing almost no recombination between a recessive mutation and male sterility in *D. simulans*/*D. mauritiana* hybrids, Coyne and Charlesworth (1986) used statistical methods to locate a sterility gene 1·1 map unit away from X-linked *forked* locus (1-56·0). Accurate to within two-tenths of a centimorgan, this represented the first mapping of a gene causing reproductive isolation in any animal species. Here we extend that study by measuring recombination between *three*

X-linked morphological markers and genes causing male sterility in hybrids of the two species pairs *D. simulans*/*D. mauritiana* and *D. simulans*/*D. sechellia*. Our intent is to determine whether "sterility genes" are closely linked to markers on the base, middle, and tip of the X chromosome, to map these genes with an eye toward cloning them in the future, and to investigate whether the locations and effects of such genes are similar in hybrids between two pairs of species.

METHODS AND MATERIALS

D. simulans, in the *D. melanogaster* group, is a cosmopolitan human commensal that probably originated in Africa (Lachaise *et al.*, 1988). *D. mauritiana* and *D. sechellia* are endemic to different islands in the Indian Ocean that lack any other members of the *melanogaster* group. The evolutionary relationship of these three species is not yet definitely known, although they are clearly more closely related to each other than to *D. melanogaster* (Lemeunier and Ashburner, 1976, 1984; Bodmer and Ashburner, 1984; Cohn *et al.*, 1984; Coyne and Kreitman, 1986). Their biogeography suggests that the two island species are independent derivatives from a *D. simulans*-like ancestor, a conclusion weakly supported by DNA sequencing of the alcohol dehydrogenase locus (Coyne and Kreitman, 1986). But DNA hybridization suggests that both island species are more closely related to each other than to *D. simulans* (Caccone *et al.*, 1988), and genetic distances among all species determined by gel electrophoresis are almost identical (Cariou, 1988). All molecular phylogenies, however, indicate that the phylogenetic branch points are very close to each other, so that the two species divergences (assuming a bifurcated tree) occurred at approximately the same time.

All three species can be crossed to yield fertile F_1 females and sterile F_1 males. The female hybrids can be backcrossed to males of either parental species, yielding backcross males that are either fertile or sterile. This segregation of male fertility permits a genetic analysis of this character.

Our methods of analysis are similar to those described by Coyne and Charlesworth (1986). Males of either *D. sechellia* or *D. mauritiana* are crossed to female *D. simulans* homozygous for a recessive X-linked marker. The heterozygous hybrid females are then backcrossed to males from the *D. simulans* marker stock, producing backcross males segregating for the marker and wild-type alleles. These males are scored for fertility; those with the mutant marker carry an X-linked segment from *D. simulans* and are usually fertile, while the wild-type males carry a segment from the island species and are largely sterile (Coyne, 1984; Coyne and Kreitman, 1986). The heterozygous, wild-type females among the backcross progeny are again backcrossed to *D. simulans* males carrying the marker. This backcross is repeated for several generations, with the two classes of males scored for fertility each generation and the heterozygous females used for the next cross. This crossing scheme with enforced heterozygosity at one locus allows a gradual reduction of linkage between that locus and X-linked sterility factors derived from the island species. The observed increase in fertility of wild-type males compared to their mutant brothers can be used to calculate the distance between the marker gene and a sterility factor (or factors) of large effect.

This crossing scheme was performed separately for three mutations at the base, middle, and tip of the X chromosome (respectively, *forked* [*f*, 1-56·0]; *miniature* [*m*, 1-36·7]; and *yellow* + *white* [*y*, 1-0·0; *w*, 4·1; these two loci were employed as a unit during crossing and scoring]. Map locations of the markers were determined in *D. simulans* by Sturtevant (1929), J. S. F. Barker (personal communication) and Coyne (unpublished); they are undoubtedly very similar in the other two species, as all three are homosequential (Lemeunier and Ashburner 1976, 1984). For each marker in *D. simulans*, repeated backcrosses were made to each of the two island species. Backcrossing between *D. simulans* and *D. mauritiana* was performed for ten generations, with males scored for fertility at generation 1, 4, 7, and 10. Backcross males between *D. simulans* and *D. sechellia* have somewhat higher fertility, and were scored in generations 1, 2, 3 and 4. Each cross was maintained at 24°C in two bottles (randomized each generation) containing Instant Drosophila Medium (Carolina Biological Supply Co.). Crosses were begun synchronously for all loci except *forked* in the *mauritiana*/*simulans* hybridization, but became somewhat asynchronous after several generations. The data for the *forked* locus in *D. simulans*/*D. mauritiana* hybrids were given in the analysis of Coyne and Charlesworth (1986) and are reproduced here for comparison.

As in our previous work, fertility is scored by observing sperm motility of hybrid males (Coyne, 1984, 1985*a*, *b*; Coyne and Charlesworth, 1986). Testes were removed from virgin males held for four days at 18°C, squashed, and inspected under a phase-contrast compound microscope. Males

Table 1 Fertility (fraction of individuals with motile sperm) of mutant and wild-type mates in backcrosses

Generation	Motile	Total	Fraction	Motile	Total	Fraction	Generation	Motile	Total	Fraction	Motile	Total	Fraction
	simulans × *mauritiana*							*simulans* × *sechellia*					
	yw			+				*yw*			+		
1	45	249	0·15	12	402	0·03	1	178	378	0·47	32	398	0·08
4	62	90	0·69	28	96	0·29	2	200	325	0·62	48	361	0·13
7	300	364	0·82	180	481	0·37	3	225	307	0·73	74	341	0·22
10	335	364	0·92	169	430	0·39	4	252	311	0·81	98	338	0·29
	m			+				*m*			+		
1	49	311	0·16	1	311	0·003	1	195	402	0·49	6	389	0·02
4	126	167	0·75	3	169	0·02	2	261	427	0·61	21	398	0·05
7	265	335	0·79	118	433	0·27	3	260	386	0·67	43	340	0·13
10	200	223	0·90	117	319	0·37	4	225	323	0·70	59	341	0·17
	f			+				*f*			+		
1	85	442	0·19	0	436	0·00	1	208	408	0·51	27	367	0·07
4	199	275	0·72	5	141	0·04	2	258	365	0·71	56	358	0·16
7	121	132	0·92	20	132	0·15	3	274	350	0·78	60	316	0·19
10	204	245	0·83	14	228	0·06	4	280	339	0·83	61	328	0·19

lacking sperm or possessing only immotile sperm were scored as "sterile", and those with at least one motile sperm as "fertile". Although males without motile sperm are invariably sterile, males with only a few motile sperm cannot always produce progeny. Scoring fertility by progeny production, on the other hand, must usually be done with much smaller samples and often confounds fertility with mating ability. Sperm motility is correlated with the true fertility of males (males with immotile or absent sperm cannot have offspring (Coyne, 1984)) and is much less laborious to score.

We describe our statistical analysis in the Results section.

RESULTS

Description of the data

In all six sets of crosses, the fertility of wild-type males is low compared to individuals carrying the markers, but gradually increases with several generations of backcrossing (table 1). These wild-type males are much more fertile in the backcross to *D. sechellia* than to *D. mauritiana*, supporting our previous observation of less genetic divergence between the former species and *D. simulans* (Coyne and Kreitman, 1986). The linkage of the wild-type allele to sterility is not equal among the markers, being highest at the *minature* locus in the *simulans/sechellia* hybridization and at the *forked* locus in the *simulans/mauritiana* hybridization.

In both hybridizations, each of the three markers along the X chromosome is linked to loci causing hybrid sterility; this results from interactions between alleles from *mauritiana* or *sechellia* and autosomal and/or Y-linked alleles from *simulans*. Each of the recessive markers is linked to the other two, however, so we cannot yet assume that there are three distinct "sterility" genes. Such a conclusion requires statistical analysis of the data, which also yields information about the map distance between the markers and any loci causing sterility. We now consider several models allowing us to estimate the location of these loci.

Model of a single X-linked locus causing complete sterility

Let the recessive marker gene under consideration be a, and its wild-type allele be a^+. Let the sterility factor closest to it be b^s, and its alternative alleles be b^f; sterility of course is only expressed on a background of the appropriate Y-linked or autosomal genes from *D. simulans*. We assume here that males carrying b^s on this background are completely sterile. The *simulans* marker X chromosome thus has the constitutution ab^f and the wild-type chromosome from the other species is a^+b^s. With the present data, the frequency of recombination between the marker and the sterility factor closest to it appears in many cases to be higher than in the case of *forked* in the *simulans/mauritiana* cross analyzed previously (see table 2).

Table 2 Estimates of recombination frequencies and their standard errors for the full sterility model

Cross:	*simulans* × *mauritiana*	*simulans* × *sechellia*
Marker	Recombination frequency	Recombination frequency
f	0·0125 ± 0·0020	0·0814 ± 0·0054
m	0·0470 ± 0·0031	0·0562 ± 0·0049
yw	0·0662 ± 0·0035	0·1037 ± 0·0063

It is therefore necessary to modify the approach used by Coyne and Charlesworth (1986) to include the possibility that some marker males are sterile because they inherited the sterility factor as a result of a crossover between the ab^f and a^+b^s chromosomes in the previous generation. The frequency of this event is non-negligible unless the recombination between these loci is very rare.

We can proceed as follows. Let p_i be the probability that an individual lacking b^s is sterile, due to the effects of the genetic background in generation i. Let q_i be the probability that a male carrying the marker in the ith generation is sterile. Let r be the frequency of recombination between the marker gene and the sterility gene in question. On the assumption of multiplicative effects on sterility (*i.e.*, a male is sterile either because of his genotype at the given locus, or because of background effects), we have

$$q_i = p_i + r(1-r)^{i-1}(1-p_i) \qquad (1)$$

since the probability that the female parent has the genotype a^+b^s/ab^f is $(1-r)^{i-1}$ and the probability that this parent produces a sterile a male progeny is r. Hence,

$$p_i = \{q_i - r(1-r)^{i-1}\}/\{1-r(1-r)^{i-1}\}. \qquad (2)$$

Similarly, the probability that an a^+ male is sterile is

$$P_i = p_i + (1-r)^i(1-p_i). \qquad (3)$$

If the number of sterile wild-type (a^+) males in generation i is n_i in a sample of size N_i, then the log-likelihood of the sample is equal to

$$\ln L = C + S\{n_i \ln P_i + (N_i - n_i)\ln(1-P_i)\}. \qquad (4)$$

Following Coyne and Charlesworth (1986), we equate q_i to the observed frequency of sterile marker males in generation i. The maximum-likelihood estimate of r can be obtained by equating $\partial \ln L/\partial r$ to zero, substituting the expression for p_i of equation (2) into equation (3).

Approximations to the variances of the estimate of r can also be obtained by the method of Coyne and Charlesworth (1986), with appropriate modifications. Details of the formulae are omitted here, but will be supplied on request. Table 2 gives the estimates of the r values and their standard errors derived from the more accurate of the two approximations to the variances for the different crosses studied here. The expected numbers of fertile a^+ males predicted from these estimates are given in table 3.

It will be seen that the estimates of the recombination frequencies are rather low; the largest estimate is 0·1 for *yw* in the *simulans* × *sechellia* cross. This suggests that there are at least three sterility genes distinguishing each pair of species, since the distances between the markers are much greater than these recombination frequencies. This conclusion must, however, be treated with some caution, for in most cases there is significant disagreement between the frequencies of sterile wild-type males predicted by the model and those observed in the backcross generations. There is good agreement between observed and expected frequencies only in the cases of the *simulans* × *sechellia* crosses involving *m* and *yw*.

To some extent, this disagreement may be due to between-generation environmental effects on fertility. In the case of *f* in the *simulans* × *mauritiana* cross, the analysis of the full data set by Coyne and Charlesworth (1986) showed that the source of the disagreement was the unusually high frequencies of fertile males of both genotypes in generation 7. Inspection of the data in table 1 does not show any effects of this kind for the other crosses, so that there is no direct evidence to support this interpretation of the discrepancies. Indeed, there is evidence for some systematic effects. For the cross of *simulans* × *mauritiana* involving *yw* and of *simulans* × *sechellia* involving *f*, there is an excess of fertile wild-type males over expectation in the early backcross generations and a deficiency in later ones. This could occur if the sterility gene was partially penetrant, so that more fertile males would be produced than expected with complete penetrance for a given recombination frequency. This effect would be strongest in the early generations when there are fewer wild-type males carrying the fertility allele. We consider the consequences of incomplete penetrance in the next section.

Table 3 Observed and expected numbers of fertile wild-type males for full sterility model

Generation	Fertile		Total
	Observed	Expected	
simulans × *mauritiana*: + vs. f^2			
1	0	1·06	436
4	5	5·06	141
7	20	10·23	132
10	14	22·66	228
$\chi^2_3 = 14{\cdot}87$ $(0{\cdot}001 < P < 0{\cdot}01)$.			
simulans × *mauritiana*: + vs. *m*			
1	1	2·42	311
4	3	23·30	169
7	118	101·64	433
10	117	112·83	319
$\chi^2_3 = 25{\cdot}03$ $(P < 0{\cdot}001)$.			
simulans × *mauritiana*: + vs. *yw*			
1	12	4·50	414
4	28	16·76	96
7	180	158·01	481
10	169	203·61	430
$\chi^2_3 = 37{\cdot}53$ $(P < 0{\cdot}001)$.			
simulans × *sechellia*: + vs. f^2			
1	27	16·59	367
2	56	42·74	358
3	60	59·76	316
4	61	83·31	328
$\chi^2_3 = 19{\cdot}52$ $(P < 0{\cdot}001)$.			
simulans × *sechellia*: + vs. *m*			
1	6	11·24	389
2	21	28·08	398
3	43	38·43	340
4	59	51·52	341
$\chi^2_3 = 6{\cdot}33$ $(0{\cdot}05 < P < 0{\cdot}10)$.			
simulans × *sechellia*: + vs. *yw*			
1	32	21·69	398
2	48	48·17	361
3	74	76·34	341
4	98	104·98	338
$\chi^2_3 = 5{\cdot}95$ $(0{\cdot}1 < P < 0{\cdot}2)$.			

Model of a single X-linked partial sterility locus

It is straightforward to modify the above model to allow for a probability $j < 1$ that a male carrying the b^s allele is sterile. Equations (2) and (3) become:

$$p_i = \{q_i - jr(1-r)^{i-1}\}/\{1 - jr(1-r)^{i-1}\} \quad (5)$$

$$P_i = p_i + j(1-r)^i(1-p_i). \quad (6)$$

Inspection of the two-dimensional log-likelihood surface generated by equations (5) and (6) showed that it was rather rugged, rendering iterative searches difficult. The location of the maximum was therefore found by inspection. In the two cases mentioned above, use of these formulae in a two-dimensional maximum likelihood scheme leads to a considerable improvement in fit over the previous case (where j is assumed equal to 1). The results for these cases are shown in table 4, where the standard errors of the estimates of r and j were obtained by the two-dimensional generalization of the second approximation method used earlier. As would be expected, the estimates of the recombination frequencies are lower than was obtained on the model of complete penetrance of the sterility gene.

Models of multiple X-linked sterility loci

The model of incomplete penetrance does not help to explain the data on *miniature* in the *simulans* × *mauritiana* cross, where there are too few fertile individuals in the early generations, rather than too many (table 3). A possible explanation of this discrepancy is that more than one sterility factor is present in the neighborhood of the marker, and that fertile males are produced only when both of these are recombined out of the a^+ chromosome. This indicates that models of more than one sterility gene must be considered. This is particularly important because we have evidence for linkage of each marker with a sterility gene, and we must consider whether these represent three distinct sterility genes or whether the linkage between the central marker *m* and sterility is actually due to one of the loci linked to *f* or *yw*.

Some qualitative insights into this question may be obtained as follows. Consider first the multiplicative model of fertility effects of different loci

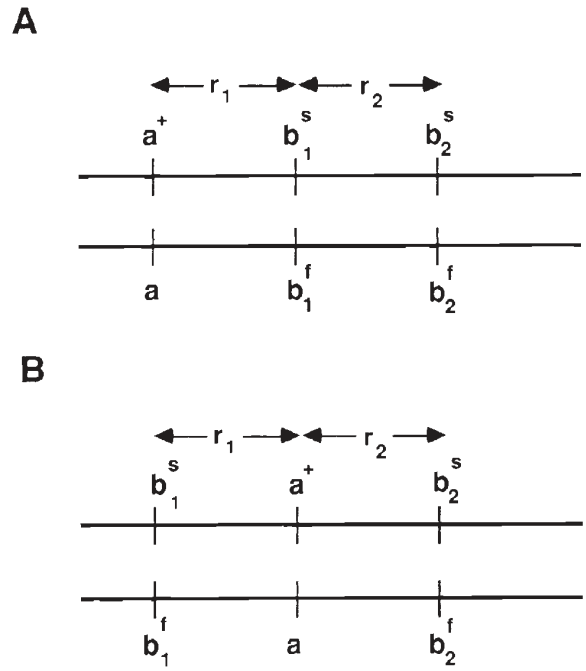

Figure 1 Possible linkage relationships between a morphological marker, a, and two sterility loci, b_1 and b_2. Recombination values = r. See text for further details. (a) Both sterility loci on the same side of the marker. (b) Sterility loci on either side of the marker.

used above for the interaction between a single X-linked locus and the background, but with two sterility loci b_1 and b_2, each located to the right (or left) of the marker (fig 1(a)). If these loci are linked closely enough that double crossovers can be ignored, then the probability that a fertile a^+ chromosome (genotype $a^+b_1^fb_2^f$) is produced in the first backcross generation is r_1, the recombination frequency between a and the closer of the two sterility genes. The other two classes of a^+ males are sterile. Once such a chromosome has been

Table 4 Observed and expected numbers of fertile wild-type males for the partial sterility model

simulans × *mauritiana*: + vs. *yw* $r = 0{\cdot}029 \pm 0{\cdot}008$ $j = 0{\cdot}73 \pm 0{\cdot}07$			*simulans* × *sechellia*: + vs. f^2 $r = 0{\cdot}026 + 0{\cdot}013$ $j = 0{\cdot}85 \pm 0{\cdot}04$		
	Fertile wild-type males			Fertile wild-type males	
Generation	Observed	Expected	Generation	Observed	Expected
1	12	19·51	1	27	32·93
4	28	23·68	2	56	50·08
7	180	163·82	3	60	57·51
10	169	183·48	4	61	65·00
	$\chi^2_3 = 8{\cdot}50$ $(0{\cdot}02 < P < 0{\cdot}05)$.			$\chi^2_3 = 2{\cdot}43$ $(0{\cdot}3 < P < 0{\cdot}5)$.	

produced, recombination has no further effect. The other recombinant a^+ chromosome (genotype $a^+b_1^sb_2^f$) produces $a^+b_1^fb_2^f$ with probability r_1 and so can be lumped with the non-recombinant $a^+b_1^sb_2^s$ chromosome as far as the production of fertile a^+ males is concerned. Thus, for the early backcross generations (when the problem of the generation of sterile *a* males by recombination can be ignored), this case behaves like the single-locus case, with the proviso that the recombination frequency that is estimated is between the marker and the closer of the two sterility genes, r_1.

The effect of two sterility genes is to *reduce* the proportion of fertile marker males below that expected on the basis of a single gene with recombination frequency r_1, since recombination in either interval will produce a sterile *a* male. This tends to inflate somewhat the estimate of the recombination frequency, although the effect will probably not be large unless the recombination frequency itself is large. Thus, the presence of two sterility genes that interact multiplicatively and are both located on the same side of the marker gene generates an estimate of *r* that slightly overestimates the recombination frequency between the marker gene and the nearer of the two loci. A similar conclusion holds if there are more than two such loci. This situation does not, therefore, lead to a serious bias in the direction of spuriously tight linkage between the marker and the sterility gene(s). It will also not explain the data on *m* in the *simulans* × *mauritiana* cross (table 3), since it behaves in essentially the same way as the single locus case.

The second situation assumes that the marker lies between the two sterility loci (fig. 1(b)). In this case, a double crossover is required to generate a fertile a^+ male by recombination between the two parental chromosomes, since he must carry a chromosome of genotype $b_1^fa^+b_2^f$. Clearly, the probability of this is very low if the loci are closely linked and if there is interference of the intensity observed in *D. melanogaster* (*e.g.*, Baker and Carpenter, 1972). However, single recombinants between the two parental chromosomes produced in one generation can experience recombination in the other interval in a later generation, thereby generating a fertile a^+ male. The frequency of these events will be given by the product of the respective recombination frequencies, and so will be much higher than the frequency of double crossovers. In this situation, one might expect a slower initial increase in the frequency of fertile a^+ males than on the single locus model, since double events are required for their production. For the same reason, if this model applies, the single locus model will underestimate the distances of the sterility genes from the marker.

The time course of appearance of fertile a^+ males can be derived as follows. As before, let r_1 and r_2 be the recombination frequencies between the marker and the two sterility loci. Let *c* be the coefficient of coincidence, such that the probability of a double crossover is cr_1r_2. This is the probability that a fertile a^+ male is produced by recombination between the two parental chromosomes. The probabilities that single recombinant chromosomes of the genotypes $b_1^fa^+b_2^s$ and $b_1^sa^+b_2^f$ are produced are $r_1(1-cr_2)$ and $r_2(1-cr_1)$ respectively. In backcross generation *i*, a sterile a^+ male will be produced either by a chromosome that has undergone no recombination in its whole history, or by a chromosome that has experienced only single crossovers in regions 1 or 2. The probability of these is simply $(1-r_1-r_2+cr_1r_2)^i$. The probability that an a^+ chromosome of generation *i* has experienced $j(j \leqq i)$ single crossovers in region 1 and no crossover in regions 2 or 1 in the remaining $i-j$ generations is

$$Q_{1ij} = \binom{i}{j} r_1^j (1-cr_2)^j (1-r_1-r_2+cr_1r_2)^{i-j}. \quad (7a)$$

Similarly, the probability that an a^+ chromosome of generation *i* has experienced $j(j \leqq i)$ single crossovers in region 2 and no crossover in region 2 or 1 in the remaining $i-j$ generations is

$$Q_{2ij} = \binom{i}{j} r_2^j (1-cr_1)^j (1-r_1-r_2+cr_1r_2)^{i-j}. \quad (7b)$$

The net probability that an a^+ chromosome of generation *i* carries one or other of the sterility genes is thus

$$Q_i = (1-r_1-r_2+cr_1r_2)^i + \sum_{j=0}^{i} (Q_{1ij}+Q_{2ij}) \quad (7c)$$

and the net probability of sterility of an a^+ male of generation *i* is thus

$$P_i = p_i + (1-p_i)Q_i. \quad (7d)$$

where p_i is the probability of sterility due to the background genotype.

If p_i is estimated directly from the frequency of sterile *a* males in generation *i*, without correcting for the introduction of sterility alleles at the *b* loci into the *a* chromosomes by recombination, we should overestimate the recombination frequencies by using equations (7), since we will overestimate p_i in equation (7d) and thereby underestimate Q_i. This should therefore provide an upper bound to the distances of the fertility genes from the

Table 5 Observed and expected numbers of fertile wild-type males for the two-locus model and the *simulans* × *mauritiana*: + vs. *m* data

$r_1 = r_2 = 0{\cdot}11$; $c = 0{\cdot}2$			$r_1 = r_2 = 0{\cdot}11$; $c = 0{\cdot}5$			$r_1 = r_2 = 0{\cdot}15$; $c = 0{\cdot}2$		
	Fertile wild-type males			Fertile wild-type males			Fertile wild-type males	
Generation	Observed	Expected	Generation	Observed	Expected	Generation	Observed	Expected
1	1	0·12	1	1	0·30	1	1	0·22
4	3	15·28	4	3	16·17	4	3	25·78
7	118	101·00	7	118	103·02	7	118	152·41
10	117	132·60	10	117	133·76	10	117	182·50
	$\chi^2_3 = 21{\cdot}09$ ($P < 0{\cdot}001$)			$\chi^2_3 = 16{\cdot}32$ ($P < 0{\cdot}0001$)			$\chi^2_3 = 54{\cdot}15$ ($P < 0{\cdot}001$)	

marker on this model. Due to its complexity, a full maximum likelihood analysis was not carried out on the data on *m* in the *simulans* × *mauritiana* cross, which is the most likely candidate for this model. Instead, assumed values of the three parameters were used to calculate the expected numbers of fertile m^+ males, and the goodness of fit determined from the χ^2 value.

We found that equal spacing of the two markers, with $r_1 = r_2 = 0{\cdot}11$, provided the best fit (table 5). High coincidence values gave the lowest χ^2 values, but a coincidence in the neighbourhood of 0·2 seems most probable for these map distances (Baker and Carpenter 1972). Even the best fits leave a highly significant disagreement, however. This is largely due to the low frequency of fertile m^+ males in generation 4, which would be hard to explain on almost any two-locus model. If the data for this generation are omitted, then the χ^2_2 value for $r_1 = r_2 = 0{\cdot}11$ and $c = 0{\cdot}5$ is reduced to 5·93 ($0{\cdot}05 < P < 0{\cdot}1$), so that an adequate fit can be obtained. Another possibility is, of course, that b^f alleles at more than two linked loci are required for fertility.

The most important conclusion, however, is that the recombination frequencies are only increased to about 10 per cent from the previous estimate of 5 per cent. Since this is in any case an overestimate, and a very significantly worse fit is obtained with recombination frequencies of 15 per cent, we can be fairly confident that the sterility loci are indeed close to *miniature* and are distinct from the loci linked to the other markers.

Another possibility with two linked loci is that their joint effects on fertility may deviate from the multiplicative model used so far. An extreme alternative is to assume that sterility only occurs in males of genotype $b_1^s b_2^s$, *i.e.*, there is duplicate gene epistasis. If the marker locus is located to one side of the sterility loci, then either class of single recombinant a^+ male will be fertile (see fig. 1). If the loci are tightly linked, then the frequency of production of fertile a^+ males by recombination between the parental chromosomes is approximately $r_1 + r_2$. The frequency of sterile *a* males produced is equal to r_1. This means that q_i in equation (1) is smaller than would be expected for a given frequency of production of fertile a^+ males, so that the estimate of recombination frequency obtained on the single gene model will lie between r_1 and $r_1 + r_2$. At any rate, this type of gene action will not underestimate the distance between the marker and the further of the two sterility loci, if the marker locus lies to one side of the sterility loci.

If the marker locus lies *between* the sterility loci, the frequency of production of fertile a^+ males by recombination between the parental chromosomes is again approximately $r_1 + r_2$, assuming tight linkage, but the frequency of sterile *a* males produced is equal to the frequency of double crossing over, and is therefore much smaller. The estimation procedure used here will therefore tend to underestimate the value of $r_1 + r_2$. The extent of this bias can be investigated by using the estimation procedure of Coyne and Charlesworth (1986), which assumes that the sterility of marker males is due entirely to effects of the genetic background. Application of this method to the present data yields estimates of the recombination frequency only slightly larger than those of table 2, and the differences are biologically unimportant (the largest being a value of 0·114 instead of 0·104 for *yw* in the *simulans* × *sechellia* cross). The conclusion that sterility loci are closely linked to each marker therefore seems robust to deviations from the simplest model.

DISCUSSION

Our statistical analysis strongly implies that at least three X-linked loci with large effects on sterility

have diverged among each of our two species pairs. Adding these results to those from previous studies of these species (Coyne 1984, 1985a; Coyne and Kreitman, 1986), we conclude that at least eight "sterility" loci have diverged between *simulans/mauritiana* and at least six between *simulans/sechellia*. As in previous studies, virtually every chromosome segment examined harbours genes causing sterility between closely-related, crossable species (Dobzhansky, 1936; Wu and Beckenbach, 1983; Coyne, 1984; Coyne and Kreitman, 1986; Orr, 1987; Naveira and Fontdevila, 1986; see Orr 1988 for an exception). Because some of these taxa are in the first stages of speciation (indeed, it is questionable whether the incompletely isolated pairs *D. simulans/D. sechellia* and *D. simulans/D. mauritiana* could remain distinct in sympatry), it is obvious that postzygotic reproductive isolation is usually caused by changes at several to many genes. There is certainly little support for the notion that "macromutations" play an important role in speciation.

The presence of several X-linked genes causing heterogametic sterility in hybrids supports the hypothesis of Charlesworth *et al.* (1987) that such sterility is a by-product of the effects of recessive or partially recessive alleles fixed by natural selection after separation of two lineages. Under this model, the rate of substitution of loci on the X exceeds that on the autosomes in random-mating populations, and so a larger number of genes on the X chromosome would distinguish the populations than for a comparable sized part of the autosomal genome. The chance fixation of slightly deleterious alleles would produce the opposite pattern, and neutral substitutions would occur at the same rate for X-linked and autosomal loci.

Because sterility effects are closely linked to three randomly-chosen morphological markers, it is likely that we have missed additional X-linked "sterility" genes. The paucity of X-linked markers in these species renders us unable to investigate this point, but such an experiment is in principle possible by observing the linkage between sterility and species-specific DNA restriction fragments on the X chromosome. In at least one pair of species there seems to be many such loci, because *no* X-linked segment from *D. serido* was compatible with fertility in a *D. buzzatii* genome (Naveira and Fontdevila, 1986). This situation probably does not obtain in *D. pseudoobscura/D. pseudoobscura bogotana* hybrids, however, because 70 per cent of the X chromosome has no effect on hybrid fertility (Orr, 1988). Nevertheless, important X-linked effects on hybrid sterility or viability have been observed in every genetic analysis of postzygotic isolation (Coyne and Orr, 1988), suggesting that the evolution of sterility and inviability involves a common pathway in all animals. In addition, because both island and non-island species show this pattern, there is little genetic support for the notion that colonizing populations speciate in novel ways (Templeton, 1981).

Because the phylogeny of these three species is still unresolved, it is formally possible that the similar findings in two species pairs do not represent two independent evolutionary events. That is, if *D. sechellia* and *D. mauritiana* are more closely related to each other than to *D. simulans*, the genetic similarities in the *D. simulans/D. sechellia* and *D. simulans/D. mauritiana* crosses may represent only two samples of the same evolutionary event—the divergence between *D. simulans* and the common ancestor of both island species. This seems unlikely, however, because all biochemical analyses show that these three species diverged at approximately the same time. There would thus be little time for substantial evolutionary change between *D. simulans* and a putative ancestor of the two island species.

Could the "sterility" genes linked to our markers be identical in both species pairs? This would imply "hot spots" for the evolution of reproductive isolation. Unfortunately, our data cannot answer this question, because our map distances require assumptions about the fertility effects of loci. We might be tempted, for example, to assert that the sterility gene linked to *forked* in the *simulans/mauritiana* cross is different from that linked to *forked* in the *simulans/sechellia* cross, because the map distance of the former ($0{\cdot}0125 \pm 0{\cdot}002$) differs significantly from that of the latter ($0{\cdot}0814 \pm 0{\cdot}005$). However, the latter gene could be much closer to *forked* if it has a less-than-complete effect on fertility (table 4).

We are now in a position to narrow down the location of such loci using DNA-based mapping, which will ultimately enable us to clone and sequence genes important in speciation.

Acknowledgements This work was supported by NIH grant GM 38462 and a grant from the Louis Block Fund of the University of Chicago to JAC, and by NSF grant BSR 8516629 to BC. We thank Allen Orr for comments.

REFERENCES

BAKER, B. S. AND CARPENTER, A. T. C. 1972. Genetic analysis of sex chromosomal meiotic mutants in *Drosophila melanogaster*. *Genetics*, *71*, 255–286.

BODMER, M. AND ASHBURNER, M. 1984. Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature*, *309*, 425-430.

CACCONE, A., AMATO, G. D. AND POWELL, J. R. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics*, *118*, 671-683.

CARIOU, M. L. 1988. Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. *Genet. Res.* (In press).

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

COHN, V. H., THOMPSON, M. A., AND MOORE, G. P. 1984. Nucleotide sequence comparison of the *Adh* gene in three Drosophilids. *J. Molec. Evol.*, *20*, 31-37.

COYNE, J. A. 1984. Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. USA*, *81*, 4444-4447.

COYNE, J. A. 1985a. The genetic basis of Haldane's rule. *Nature*, *314*, 736-738.

COYNE, J. A. 1985b. Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. *Genet. Res. 46*, 169-192.

COYNE, J. A. AND CHARLESWORTH, B. 1986. Location of an X-linked factor causing sterility in male hybrids of *Drosophila simulans* and *D. mauritiana*. *Heredity*, *57*, 243-246.

COYNE, J. A. AND KREITMAN, M. 1986. Evolutionary genetics of two sibling species of *Drosophila*, *D. simulans* and *D. mauritiana*. *Evolution*, *40*, 673-691.

COYNE, J. A. AND ORR, H. A. 1988. Two rules of speciation. In Endler, J. and Otte, D. (eds) *Speciation and its Consequences*, Sinauer (In press).

DOBZHANSKY, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, *21*, 113-135.

GOULD, S. J. 1980. Is a new and general theory of evolution emerging? *Paleobiology 6*, 119-130.

HALDANE, J. B. S. 1922. Sex-ratio and unisexual sterility in hybrid animals. *J. Genet.*, *12*, 101-109.

LACHAISE, D., CARIOU, M.-L., DAVID, J. R., LEMEUNIER, F., AND TSACAS, L. 1988. The origin and dispersal of the *Drosophila melanogaster* subgroup: a speculative paleobiogeographic essay. *Evolutionary Biology* (In press).

LANCEFIELD, D. E. 1929. A genetic study of crosses of two races or physiological species of *Drosophila obscura*. *Z. indukt. Abstamm.-u. Vererbungslehre*, *52*, 287-317.

LEMEUNIER, F. AND ASHBURNER, M. 1976. Relationships within the *melanogaster* subgroup of the genus *Drosophila* (Sophophora). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. *Proc. R. Soc. London B*, *193*, 275-294.

LEMEUNIER, F. AND ASHBURNER, M. 1984. Relationships within the *melanogaster* subgroup of the genus *Drosophila* (Sophophora). II. Phylogenetic relationships between six species based upon polytene chromosome banding. *Proc. R. Soc. London B*, *193*, 275-294.

NAVEIRA, H. AND FONTDEVILA, A. 1986. The evolutionary history of *Drosophila buzzatti*. XII. The genetic basis of sterility in hybrids between *D. buzzatii* and its sibling *D. serido* from Argentina. *Genetics*, *114*, 841-857.

ORR, H. A. 1987. Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimillis*. *Genetics*, *116*, 555-563.

ORR, H. A. 1988. Genetics of sterility in hybrids between two subspecies of *Drosophila*. *Evolution* (In press).

STURTEVANT, A. H. 1929. The genetics of *Drosophila simulans*. *Carnegie Inst. Wash. Pub. 399*, 1-62.

TEMPLETON, A. R. 1981. Mechanisms of speciation—a population genetic approach. *Ann. Rev. Ecol. Syst*, *12*, 23-48.

WU, C.-I., AND BECKENBACH, A. T. 1983. Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *D. persimilis* and identification of hybrid sterility factors. *Genetics*, *105*, 71-86.