

# Location of an X-linked factor causing sterility in male hybrids of *Drosophila simulans* and *D. mauritiana*

Jerry A. Coyne\* and  
Brian Charlesworth†

\* Department of Zoology, The University of  
Maryland, College Park, MD 20742, USA

† Department of Biology, The University of Chicago,  
Chicago, IL 60637, U.S.A.

**We report the first mapping of a genetic factor responsible for reproductive isolation: a small segment of genome strongly affecting sperm motility in hybrids between the sibling species *Drosophila simulans* and *D. mauritiana*. Maximum-likelihood analysis of data from ten generations of backcrossing places this factor at  $1.1 \pm 0.2$  map units from the *forked* locus, at position  $54.9 \pm 0.2$  or  $57.1 \pm 0.2$  on the X chromosome.**

## INTRODUCTION

A complete picture of speciation would require examination of individual loci causing reproductive isolation, determination of their normal function, and explanation of their evolutionary change. The first step in this process is the location of genes with important effects on interspecific isolation. Although there have been several attempts to estimate the number of genes affecting such characters (e.g., Dobzhansky, 1936; Coyne, 1984), the nature of the genetic analysis usually precludes precise mapping of loci. In this study we provide what we believe is the first location of a gene with profound effects on hybrid sterility.

*Drosophila simulans* and *D. mauritiana* are sibling species, the former cosmopolitan and the latter restricted to the volcanic island of Mauritius, 1000 km east of Madagascar (Tsacas and David, 1974). The close relationship of these species is shown by chromosome banding patterns (Lemeunier and Ashburner, 1976), gel electrophoresis (Gonzales *et al.*, 1982; Ohnishi *et al.*, 1983), and DNA sequencing (Bodmer and Ashburner, 1984; Cohn *et al.*, 1984). *D. mauritiana* appears from biogeographic evidence to have arisen from a population of proto-*simulans* colonizing Mauritius.

The cross between *D. simulans* females and *D. mauritiana* males produces sterile male and fertile female offspring. (The reciprocal cross is more difficult but gives identical results). When the hybrid females are back-crossed to either parent,

a few fertile males appear among the progeny (David *et al.*, 1976). Using recessive alleles as markers of the *D. simulans* genome, Coyne (1984) showed that at least five loci were responsible for male sterility in this backcross. The largest effect mapped to the X chromosome and was closely linked to the *forked* allele used as a marker. This fortuitous linkage between a sterility factor and a morphological mutant permits us to map the factor by measuring the decay of the association over several generations. More traditional methods of mapping are impractical because few mutants exist in these species.

## METHODS AND MATERIALS

Females from a stock of *D. simulans* homozygous for the bristle mutant *forked* (*f*; 1–56.0) were crossed to males from a strain of *D. mauritiana* provided by the Mid-America *Drosophila* Stock Center, Bowling Green, Ohio. The heterozygous hybrid females were backcrossed to *forked* *D. simulans* males, and the *forked* and wild-type (*f*<sup>+</sup>) male progeny scored for fertility. Sixty heterozygous non-*forked* (*f*/*f*<sup>+</sup>) females from the first backcross were again backcrossed to *forked* *D. simulans* males. This cross was repeated for 10 backcross generations, with males scored each generation and heterozygous females used for further crossing. This scheme allows gradual recombinational decay of linkage between the *mauritiana*-derived *f*<sup>+</sup> allele and any sterility factors. All

crosses were maintained at 24° in two bottles (randomized each generation) containing Carolina Instant *Drosophila* Medium.

The measurement of fertility is identical to that used in our previous work (Coyne, 1984, 1985): after virgin males were held for four days at 18°C, their testes were dissected in Ringer's solution and examined under a phase-contrast compound microscope. Males completely lacking motile sperm were scored as "nonmotile", and those with one or more motile sperm as "motile". Although males in this latter class are sometimes sterile, the method permits rapid scoring of large samples and provides reliable statistical discrimination of the effects of chromosomes (Coyne, 1984). The analysis of our data is described in the following section.

## RESULTS AND DISCUSSION

Table 1 gives the number of males with and without motile sperm in the *f* and *f*<sup>+</sup> classes over ten generations. We confirm the large difference in fertility between these genotypes in the first generation (Coyne, 1984), and show that the *f*<sup>+</sup> genotype is always much more sterile than *f* during the course of the study. This indicates tight linkage between the *forked* locus and a fertility factor. The fertility of the *f*<sup>+</sup> class gradually rises as the linkage decays, but the fertility of the *forked* males increases more quickly as the autosomal DNA from *D. mauritiana* is replaced by that from *D.*

*simulans*. The increase in fertility of both genotypes is somewhat irregular after the first generation, particularly in generation 7 when fertility is anomalously high in both classes. This probably reflects environmental factors.

### Analysis of the data

We examine a model in which a backcross male is assumed to be sterile if he possesses a *mauritiana* allele at an X-linked locus close to *f*, regardless of his constitution with respect to autosomal loci. (This is primarily due to incompatibility between the *mauritiana* X chromosome and the *simulans* Y chromosome [Coyne, 1985]). He may also be sterile because of his Y-chromosome or autosomal ("background") constitution, as shown by the occurrence of a substantial frequency of nonmotile *f* backcross males even in the last generations of backcrossing. There is also a contribution of a distally-located X chromosome motility factor to this category, as discussed below. Since the frequency of sterile *f* males is usually much higher than the frequency of fertile *f*<sup>+</sup> males, we shall neglect the possible contribution of crossing over between *f* and the putative sterility locus to the class of sterile *f* males. A more elaborate analysis that takes this contribution into account yields almost identical results, and will not be presented here.

The effects of the different loci on fertility are assumed to be multiplicative, i.e., a male is sterile

**Table 1** Number of males with motile or nonmotile sperm in *forked* and wild-type classes over ten generations of backcrossing. The expected values in the wild-type class predicted from the theoretical model are given in parentheses

Backcross generation	Forked			Wild-type ( <i>f</i> <sup>+</sup> )		
	Motile	Nonmotile	Fraction motile	Motile	Nonmotile	Fraction motile
1	85	357	0.192	0 (1.0)	436 (435.0)	0.000
2	42	46	0.477	0 (1.2)	107 (105.8)	0.000
3	105	47	0.691	2 (3.1)	124 (122.9)	0.016
4	199	76	0.724	5 (4.8)	136 (136.2)	0.035
5	53	14	0.791	5 (2.8)	55 (57.2)	0.083
6	177	47	0.790	9 (8.9)	153 (153.1)	0.058
7	121	12	0.910	20 (9.7)	112 (122.3)	0.151
8	181	60	0.751	10 (12.8)	177 (174.2)	0.053
9	224	47	0.823	23 (22.3)	241 (241.7)	0.087
10	204	41	0.832	14 (21.5)	214 (206.5)	0.061

Estimated recombination value: 0.0119

First approximation to variance:  $8.6820 \times 10^{-7}$

Second approximation to variance:  $8.6838 \times 10^{-7}$

Standard deviation of recombination value: 0.00093

Goodness of fit:  $\chi^2$  (9 d.f.) = 20.0,  $0.02 < P < 0.05$

because of either his X-linked constitution, his background constitution, or both. If  $p_i$  is the probability of sterility due to the background in generation  $i$ , and  $r$  is the frequency of recombination between the locus of  $f$  and the sterility locus, the probability of sterility for a male of generation  $i$  is given by

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

Let  $n_i$  and  $m_i$  be the numbers of sterile and fertile  $f^+$  males in generation  $i$ : The likelihood of the set of observations from generation 1 to  $k$  is thus

$$L = \prod_{i=1}^k \binom{N_i}{n_i} P_i^{n_i} (1 - P_i)^{m_i} \quad (2)$$

where  $N_i$  is the total number of  $f^+$  males in generation  $i$ .

If the parameters  $p_i$  are replaced by the observed proportions of sterile  $f$  males in the corresponding generations,  $r$  can be estimated by finding the value which maximizes the log-likelihood obtained from eq. (2). In view of the large numbers of  $f$  males scored in most generations, the error in this procedure, compared with the more tedious process of a joint maximum-likelihood estimation of  $r$  and the  $P_i$  from the combined data on both classes of male, should be unimportant (see below). A BASIC computer program was written to obtain this estimate  $r$  by Newton-Raphson iteration of the equation  $\partial \log L / \partial r = 0$ .

An approximate estimate of the sampling variance of the estimate of  $r$  can be found from the inverse of the expectation of  $-\partial^2 \log L / \partial r^2$  (e.g. Kempthorne 1957, p. 172). This provides the first approximation to the variance of  $r$  in table 1. The contribution of the sampling errors of the  $P_i$  to the variance of  $r$  are not taken into account by this. An approximate formula including their effect was obtained by using implicit differentiation of the equation  $\partial \log L / \partial r = 0$  to calculate  $\partial r / \partial p_i$ . The variances in the estimates of the  $p_i$  were calculated from the usual binomial formula, and their contribution to the variance of  $r$  found by weighting by  $(\partial r / \partial p_i)^2$  and summing according to the delta-method rule (c.f. Bulmer 1980, p. 82). This provides the second approximation to the variance of  $r$  in table 1. The difference between the two variance approximations is small, in accordance with the large sample sizes for  $f$  males. The standard deviation of  $r$  is found from the larger variance estimate.

The goodness of fit of the model was tested by calculating expected proportions of sterile and fertile males using eq. (1) and the estimate of  $r$ , and

applying a  $\chi^2$  test (table 1). Because the  $p_i$  are estimates, not parameters, the  $\chi^2$  value is probably upwardly biased, and it is only marginally significant in any case.

The main contribution to  $\chi^2$  comes from generation 7, in which there is an excess of motile males. It can be seen from table 1 that there is an unusually low frequency of sterile  $f$  males in this generation, suggesting an environmental effect on fertility. If generation 7 is omitted from the analysis, the estimate of  $r$  becomes 0.010, with standard deviation 0.009. This is probably a more accurate estimate of  $r$ . The corresponding  $\chi^2$  is 7.58 (8 d.f.,  $p > 0.3$ ). All in all, there is generally good agreement between the observed results and the model.

A previous single-generation backcross experiment used a *simulans* X chromosome marked with  $y$  (map position 0.0),  $w$  (4.1), as well as  $f$  (56.0) (Coyne and Kreitman, 1986). This showed that there was a factor sufficiently distal to  $f$  to cause a 94.5 per cent chance of nonmotility for  $++f$  recombinant backcross males, compared with a chance of 67.9% for  $ywf$  nonrecombinants. Using a model of the interaction of this motility factor with the autosomal background similar to that employed above (such that  $++f$  male is nonmotile if he carries either the distal X factor or a background factor), the frequency of nonmotile  $++f$  males due to the introduction of the distal X factor by crossing over is estimated to be  $1 - 0.055 / 0.321 = 0.829$ . The probability of a crossover between the locus of  $w$  and the distal motility factor, for a  $++f$  chromosome, is thus estimated to be 0.171 and so the distance of the factor from  $w$  is  $0.171 \times 51.9 = 8.9$  where 51.9 is the distance from  $w$  to  $f$ . This analysis assumes full penetrance of the motility factor; if males carrying the allele conferring nonmotility were sometimes scored as motile, linkage to  $w$  would be even tighter than suggested by this estimate.

It is therefore clear that the distal X chromosome factor is very loosely linked to the proximal factor considered above. It is therefore legitimate to treat its contribution to the nonmotility of the backcross males in the same way as the autosomal background.

The large samples used in this study allow us to locate the fertility factor  $1.1 \pm 0.2$  (2 S.E.) map units from *forked*, itself mapped by Sturtevant (1929) to 56.0 on the X chromosome. Neglecting any error in the map position of *forked*, we may place our sterility factor at either  $54.9 \pm 0.2$  or  $57.1 \pm 0.2$ . (We cannot, of course, rule out the possibility that this small region contains several very tightly linked genes affecting fertility). The

only comparable result appears to be Pickering's (1983) mapping of a locus causing incompatibility between two strains of *Hordeum* at  $11.2 \pm 3.2$  map units from a locus conferring DDT resistance (the map position of the latter gene on the seventh chromosome was not determined). Such information may one day allow us to isolate the DNA from "sterility" loci, examine its normal function, and begin to understand the genetical causes of speciation.

*Acknowledgments* This work was partially supported by National Institutes of Health grant GM 32221 and National Science Foundation grant BSR-18558 to JAC, and by a grant from the Louis Block Fund of the University of Chicago to BC. We thank Allen Orr for helpful comments.

## REFERENCES

- BODMER, M. AND ASHBURNER, M. 1984. Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature*, **309**, 425-429.
- BULMER, M. G. 1980. *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, Oxford.
- COHN, V. H., THOMPSON, M. A. AND MOORE, G. P. 1984. Nucleotide sequence comparison of the Adh gene in three drosophilids. *J. Molecular Biol.*, **20**, 31-37.
- COYNE, J.A. 1984. Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. *Proc. Nat. Acad. Sci. USA*, **81**, 4444-4447.
- COYNE, J. A. 1985. The genetic basis of Haldane's rule. *Nature*, **314**, 736-738.
- COYNE, J. A. AND KREITMAN, M. 1986. Evolutionary genetics of two sibling species of *Drosophila*. *Evolution* (in press).
- DAVID, J., BOCQUET, C., LEMEUNIER, F. AND TSACAS, L. 1976. Persistence of male sterility in strains issued from hybrids between two sibling species: *Drosophila simulans* and *D. mauritiana*. *J. Genetics*, **62**, 93-100.
- DOBZHANSKY, T. 1936. Studies in hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, **21**, 113-135.
- GONZALES, A. M., CABRERA, V. M., LARRUNGA, J. M. AND GULLON, A. 1982. Genetic distance in the sibling species *Drosophila melanogaster*, *D. simulans*, and *D. mauritiana*. *Evolution*, **36**, 517-522.
- KEMPTHORNE, O. 1957. *Introduction to Genetic Statistics*. John Wiley & Sons, New York.
- 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. Roy. Soc. Lond. B*, **193**, 275-294.
- OHNISHI, S., KAWANISHI, M. AND WATANABE, T. K. 1983. Biochemical phylogenies of *Drosophila*: protein differences detected by two-dimensional electrophoresis. *Genetica*, **61**, 55-63.
- PICKERING, R. A. 1983. The location of a gene for incompatibility between *Hordeum vulgare* L. and *H. bulbosum* L. *Heredity*, **51**, 455-459.
- STURTEVANT, A. H. 1929. The genetics of *Drosophila simulans*. pp. 1-62 in Carnegie Institute of Washington Publication 399.
- TSACAS, L. AND DAVID, J. 1974. *Drosophila mauritiana* n. sp. du groupe *melanogaster* de l'île Maurice. *Bull. Soc. Ent. Fr.*, **79**, 42-46.