Genetic Dissection of Hybrid Incompatibilities Between *Drosophila simulans* and *D. mauritiana*. I. Differential Accumulation of Hybrid Male Sterility Effects on the *X* and Autosomes

Yun Tao,*,†,[1] Sining Chen,‡ Daniel L. Hartl† and Cathy C. Laurie*,[2]

**DCMB and Department of Zoology, Duke University, Durham, North Carolina 27708, ‡Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708 and †Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138*

Manuscript received January 22, 2003
Accepted for publication March 21, 2003

ABSTRACT

The genetic basis of hybrid incompatibility in crosses between *Drosophila mauritiana* and *D. simulans* was investigated to gain insight into the evolutionary mechanisms of speciation. In this study, segments of the *D. mauritiana* third chromosome were introgressed into a *D. simulans* genetic background and tested as homozygotes for viability, male fertility, and female fertility. The entire third chromosome was covered with partially overlapping segments. Many segments were male sterile, while none were female sterile or lethal, confirming previous reports of the rapid evolution of hybrid male sterility (HMS). A statistical model was developed to quantify the HMS accumulation. In comparison with previous work on the *X* chromosome, we estimate that the *X* has ~2.5 times the density of HMS factors as the autosomes. We also estimate that the whole genome contains ~15 HMS "equivalents"—*i.e.*, 15 times the minimum number of incompatibility factors necessary to cause complete sterility. Although some caveats for the quantitative estimate of a 2.5-fold density difference are described, this study supports the notion that the *X* chromosome plays a special role in the evolution of reproductive isolation. Possible mechanisms of a "large *X*" effect include selective fixation of new mutations that are recessive or partially recessive and the evolution of sex-ratio distortion systems.

POPULATION genetics theory predicts that sex chromosomes (*X* or *Y*) have unique evolutionary dynamics that differ from the autosomes (Haldane 1924). For example, newly arisen *X*-linked mutations have a greater substitution rate under selection than autosomal mutations have, provided they are recessive (or partially recessive) and favorable to the heterogametic sex (Charlesworth *et al.* 1987). The higher substitution rate of new recessives that favor the heterogametic sex may hold even for "sexually antagonistic" mutations that are detrimental to the homogametic sex (Rice 1984). However, the opposite result is obtained by considering adaptation to a sudden environmental change in which mutations that were previously deleterious and held in mutation-selection balance become beneficial and fixed by selection. In this case, *X*-linked genes evolve more slowly than autosomal genes, regardless of dominance (Orr and Betancourt 2001).

If *X*-linked loci evolve faster than autosomal loci, interspecific trait differences should map disproportionately to the *X* chromosome (*i.e.*, a "large *X*" effect). At one time, it was argued that the faster evolution of the *X* chromosome is a general pattern in Drosophila. "The genes having the greatest effect on hybrid sterility and inviability are *X*-linked" (Coyne and Orr 1989, p. 181). This suggestion of a large *X* arose from classic backcross analyses of hybrid incompatibilities (Dobzhansky 1936). In all case studies of this kind that followed, the *X* chromosome shows a disproportionate contribution to hybrid incompatibility. However, as pointed out by Wu and Davis (1993), the large *X* effect could be an observational artifact in this type of experiment due to the hemizygosity of the *X* chromosome. Although the *X* chromosome appears to be hosting a disproportionately large amount of hybrid incompatibility, the difference may actually be due to the masking of the recessive incompatibility factors on autosomes. Meaningful comparison of the contributions of the *X* and autosomes requires an experimental design that removes the confounding effect of dominance and provides sufficient resolution to detect quantitative differences in the density of genetic factors affecting the trait.

As a paradigm of studying the genetics of speciation, the *X* chromosome in *Drosophila simulans* and *D. mauritiana* has been intensively mapped for hybrid male sterility (HMS). Employing an introgression approach, Coyne and Charlesworth (1986, 1989) localized three "major" factors, each closely linked to one visible marker. A major factor is defined as a single locus or a cluster of closely linked loci ("minor" factors) that can cause

[1]*Corresponding author:* Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave., Cambridge, MA 02138. E-mail: ytao@oeb.harvard.edu

[2]*Present address:* Monsanto Company, 800 N. Lindbergh Blvd., N2NA, St. Louis, MO 63167.

complete sterility. With additional markers, Wu and coworkers covered the three regions surrounding these major factors (~60% of the *X* chromosome) by more introgressions (Perez *et al.* 1993; Cabot *et al.* 1994; Perez and Wu 1995; Davis and Wu 1996). They carefully analyzed a smaller coverage (~40% of the *X*) and mapped nine minor factors (reviewed in Wu and Hollocher 1998; two factors mapped in 9A–10E are not taken into account due to the relative lower resolution in this region).

Recent genetic analyses of autosomal sterility factors in *D. simulans* and *D. mauritiana* hybrids have removed the complications due to dominance by comparing homozygous introgressions on the autosomes with hemizygous introgressions on the *X*. With this design, Hollocher and Wu (1996) suggested a similar density of HMS factors on the *X* and the second chromosomes, while True *et al.* (1996b) inferred (qualitatively) a higher density on the *X* than on the two major autosomes. Unfortunately, the genetic resolution of these studies was not fine enough to give a reliable conclusion.

Here we report an introgression experiment designed to finely map HMS factors on the third chromosome with a resolution similar to that on the *X* chromosome. The basic idea of constructing the introgression lines follows that of True *et al.* (1996b) by tracking $P[w^+]$-tagged *D. mauritiana* fragments in repeated backcrossings to *D. simulans*. However, several improvements have been made to the original scheme. We were able to estimate the lengths of introgressed *D. mauritiana* fragments by using 38 allele-specific oligonucleotide (ASO) markers (Liu *et al.* 1996; Zeng *et al.* 2000), as well as the 38 $P[w^+]$ inserts themselves. The ASO markers were used to identify and produce a set of introgressions that covered the entire chromosome, avoiding a potential problem in the original scheme where selection during backcrossing might have biased against the recovery of introgressions with heterozygous effects in females. Fertility and viability effects of the introgressed segments were carefully assayed as homozygotes in a manner that permits comparison to previously published work on the *X* chromosome. A statistical model was developed and used to estimate the ratio of densities of HMS factors on the *X* and third chromosomes as 2.5:1. This result strongly supports the conclusion that the *X* chromosome plays a special role in the evolution of hybrid incompatibility.

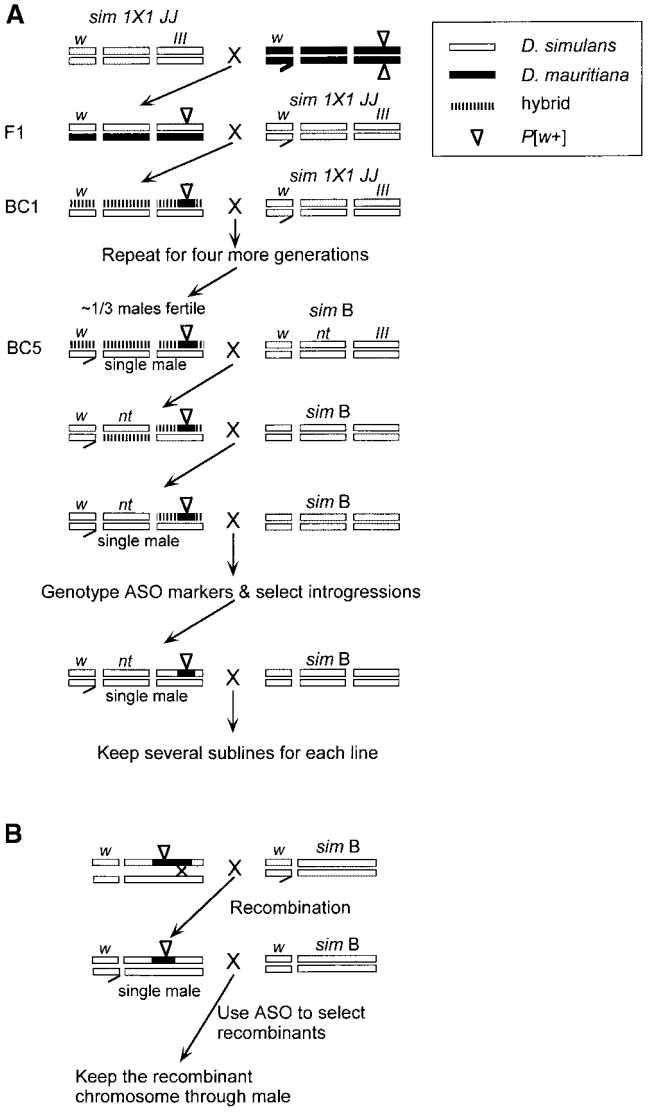

Figure 1.—Introgression scheme. The *X* and the third chromosomes (*X*, 2, and 3) are represented by short and long bars, respectively (open, *D. simulans*; solid, *D. mauritiana*; shaded, hybrid, parental origin unclear), and the *Y* chromosome by hooked short bars. The stock *sim*B has the *X* and the second chromosomes marked by *w* and *nt*, respectively. *III* represents the highly inbred third chromosome from the stock *sim 1* × *1 JJ*. (A) Stage I. Introgression lines were constructed with an isogenic background. The visible markers on the *X* and second chromosomes, and the ASO markers on the third chromosomes, were used to select introgression lines with an isogenic background. Several sublines were made for each $P[w^+]$-insert and each subline has only one introgressed segment. (B) Stage II. Shorter introgressions were made. Marker-assisted selection was used to pick a set of predetermined introgressions. See text for details. Overall, 38 $P[w^+]$-insert stocks were used to generate 259 introgression sublines. The second chromosome is not shown.

MATERIALS AND METHODS

Experimental design: A two-stage protocol was used to introgress fragments of the *D. mauritiana* third chromosome into a *D. simulans* genetic background (Figure 1). A salient feature of these schemes is to follow the introgression by tagging the *D. mauritiana* segments with $P[w^+]$-inserts, which are semidominant markers with various eye colors (from pale orange to deep red) because of position effects. Usually, the homozygous $P[w^+]$ flies, with two copies of $P[w^+]$-inserts, have darker eyes than heterozygous flies. In most cases, the eye color differences are clear-cut. For convenience, the heterozygous $P[w^+]$ fly is called "orange eyed" and the homozygous fly is called "red eyed" throughout the text.

In stage I, females of *sim 1* × *1 JJ*, a highly inbred *D. simulans*

stock (Liu *et al.* 1996), were crossed to males of $P[w^+]$-insert *D. mauritiana* males. These $P[w^+]$-insert strains were described previously (True *et al.* 1996a). Orange-eyed female progeny were then backcrossed to *sim 1* × *1 JJ* for five generations [backcross (BC)5], by which time approximately one-third of hybrid males were fertile (David *et al.* 1976). For each line, 20–40 orange-eyed males from BC5 were singly crossed to the females of *sim*B, a *D. simulans* stock visibly marked on the *X* and the second chromosomes. Using the visible and ASO markers, several sublines for each $P[w^+]$ insert were selected. All sublines were isogenic for the whole genome except one single introgressed *D. mauritiana* segment surrounding the $P[w^+]$-insert. To maintain these sublines, the hybrid chromosomes were transmitted only through males (Figure 1A).

In stage II, shorter introgressions were made. Most of the sublines from stage I had relatively long (greater than one-third of the whole chromosome) *D. mauritiana* segments, which might host many HMS factors. For high-resolution mapping, much shorter *D. mauritiana* segments were developed. From the cross in Figure 1B, 100 or more orange-eyed males were singly mated to *sim*B females. After sublines had been established, recombinants yielding shorter introgressions were identified by genotyping the males with appropriate ASO markers.

A total of 259 introgressions (28 from stage I and 231 from stage II) were selected and kept as the set of lines for various phenotypic assays. Each line was given a serial number in the format ##.##, where the number preceding the dot represents the $P[w^+]$-insert and the number following the dot represents one of several sublines tagged by the same $P[w^+]$. The cytological positions of all $P[w^+]$-inserts used are summarized in the Figure 4 legend. All lines but one (20.11) have just one uninterrupted segment of *D. mauritiana* introgressed.

Heterozygous males from all introgression lines were fertile. To remove the confounding effects of dominance, the fertility of each autosomal introgression was assayed in homozygous condition (Figure 2A). Red-eyed homozygotes were produced by crossing orange-eyed males and females of the same line. For each line tested, at least 10 red-eyed males or females were assayed for fertility. Fertility was assayed in a mating test by crossing red-eyed flies of one sex with the tester stock *D. simulans w*; *e*, where *e* (*ebony*) was used to detect nonvirgins. For each tested individual, the flanking ASO markers on maternally derived *D. mauritiana* segments were checked to exclude recombinants.

During the introgression process and the later stock maintenance, spontaneous lethal and sterile mutations may have accumulated. Because the introgression chromosomes were transmitted through the nonrecombining male, lethal and sterile mutations may significantly accumulate through Muller's ratchet. Assuming that the lethal mutation rate is ~0.005 per chromosome per generation (Simmons and Crow 1977), the frequency of lethal chromosomes is expected to be $\sim 1 - (1 - 0.005)^{20} = 9.5\%$ in a stock maintained for 1 year (~20 generations/year). If the sterile mutation rate is ~15% of that of lethals (True *et al.* 1996b), then the frequency of spontaneous sterility should be ~1.5%. Both amounts are sufficiently high that they cannot be ignored. Thus introgressed lines tested as sterile or lethal by the method in Figure 2A could be due to spontaneous mutations. To address this issue, we retested the sterile and lethal lines by producing *trans*-heterozygotes from two independently derived lines (tagged by different $P[w^+]$-inserts), usually with one shorter introgression nested within the other longer one (Figure 2B). Similar to Figure 2A, the maternal chromosome was checked for ASO markers to exclude recombinants. Introgressions that were lethal or sterile when homozygous, but not when *trans*-heterozygous, in this test are assumed to carry spontaneous mutations that arose during the course of the experiment. This method tests for mutations in the vicinity of the *P*-insert and not throughout the entire chromosome.

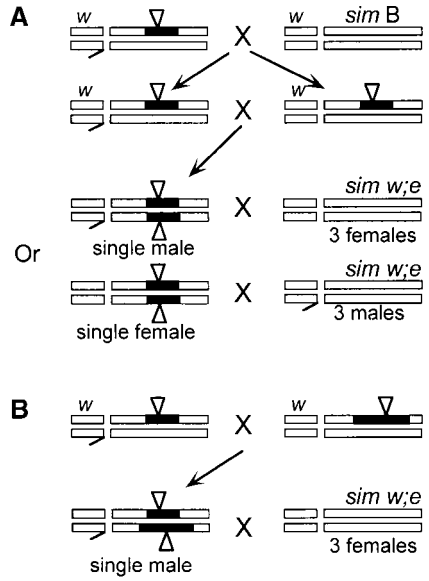

Figure 2.—Testing male fertility of introgression lines. (A) Homozygous introgression males and females were assayed for fertility by mating tests. (B) For the lines tested in A as sterile or lethal, a second-round test was carried out to make sure that sterility or lethality was not caused by spontaneous mutations. The genuine hybrid sterility/lethality should still be observed if testing *trans*-heterozygous males or females produced by crossing these lines *inter se*. The two parental lines for each *trans*-heterozygote were independently derived and were tagged with different $P[w^+]$. Usually, one parental line had a longer introgressed segment that covered the shorter one from the other parental line. Introgressions through females were checked by ASO markers to exclude recombinants. The second chromosome is not shown. Also see the Figure 1 legend.

Drosophila stocks:

D. mauritiana-1: *w* (*white*, 1–1.5 on the *D. melanogaster* map) is from J. Coyne; this stock was the progenitor for constructing the $P[w^+]$-insert lines (True *et al.* 1996a).

D. mauritiana-2: *w*; $P[w^+]$, 38 lines of independent inserts on the third chromosome (True *et al.* 1996a, Table A1); for convenience, we renumbered the lines (see Figure 4 legend).

D. simulans-1, *13w* (*white,* 1-4.1); *D. simulans*-2, *nt* (*net*, 2-0); *D. simulans*-3, *Ubx/D* (*Ultrabithorax* 3-71, *Delta* 3-64); *D. simulans*-4, *w*; *e* (*ebony*, 3-72): These stocks were also provided by J. Coyne. The descriptions of the mutations can be found in Lindsley and Zimm (1992).

D. simulans-5: *13w 1* × *1 JJ* is a line constructed by sib-pair mating for 20 generations from *13w* (Liu *et al.* 1996).

Construction of D. simulans simB: D. simulans is a very polymorphic species in its DNA sequence and so is *D. mauritiana*, but to a lesser degree (Hey and Kliman 1993). Due to their recent common ancestry, they share some alleles. To develop molecular markers, it is necessary to make both species fixed for alleles as much as possible. For this purpose, the highly inbred stock *13w 1* × *1 JJ* was used. However, this stock is

extremely weak. To remedy this problem, we constructed a stock with the third chromosome homozygous and isogenic to that of *13w 1 × 1 JJ*, but with the first and second chromosomes potentially heterozygous and derived from several different stocks (*D. simulans* stocks 1–3 described above). We first obtained female *w*/*w*; *nt*/+; *D*/+, with the notable characteristic that *nt* becomes dominant in the presence of *D*. We then crossed this female to *13w 1 × 1 JJ* males and backcrossed the male progeny *w*; *nt*/*II*; *D*/*III* to females of *13w 1 × 1 JJ*. The stock *sim*B (*w*; *nt*; *III*) was then obtained by pair mating *w*; *nt*/*II*; *III*. Here, *II* and *III* represent the isogenic second and third chromosomes from *13w 1 × 1 JJ*. The fourth chromosome was not followed.

Purging of standing lethal and sterile mutations: It is possible that some standing lethal and sterile mutations in the $P[w^+]$-tagged *D. mauritiana* chromosome could be introgressed into the *D. simulans* background. To avoid this problem, we performed a pseudo-extraction to purge any sterile and lethal mutations in the vicinity of the $P[w^+]$-insert. Because of the lack of balancers in *D. mauritiana*, we can make a segment around the $P[w^+]$-insert isogenic in a statistical sense only. Males from $P[w^+]$-insert stocks were crossed to *w*. Single F_1 males were then backcrossed to *w*. F_2 males and females heterozygous for the $P[w^+]$-insert from the same line were crossed. The three genotypes ($P[w^+]/P[w^+]$, $P[w^+]/+$, and $+/+$) were counted in the F_3 progeny. Five pair matings of homozygous $P[w^+]$-insert ($P[w^+]/P[w^+]$) were set up and their progeny (F_4) counted. Homozygous $P[w^+]$ individuals in F_3 would be less than expected (25%) if there are linked lethal mutations. Similarly, the number of F_4 progeny would be smaller if male and/or female sterile mutation(s) are nearby. For each $P[w^+]$ line, 5–10 third chromosomes were extracted. Among the 304 chromosomes extracted, 17 (5.6%) chromosomes (with 14 different $P[w^+]$-inserts) were found to be subvital or lethal. Two chromosomes (0.7%) did not produce any viable homozygous $P[w^+]$ flies at all. Three chromosomes (1%) were detected with sterile mutations. The standing lethal and sterile mutations in the $P[w^+]$-insert stocks were therefore ~0.01 per chromosome. For each $P[w^+]$-insert line, one extracted chromosome line with a combined maximum of viability and fertility was selected for later experiments. All flies were raised with standard cornmeal molasses agar medium at room temperature (21°–24°).

Molecular markers: ASO markers (Saiki *et al.* 1986) were used throughout this study. The ASO probes are pairs of 15-mers with the same sequences but one or more species-specific substitutions. Each member of the pair can hybridize specifically to DNA of one species but not the other. The ASO probes were developed using the following steps:

1. From FlyBase (http://flybase.bio.indiana.edu/), known *D. melanogaster* gene sequences in a particular cytological region were identified. For the most part, exonic sequences flanking intron(s) were used for designing PCR primers.
2. Two stocks, *w* of *D. mauritiana* and *13w 1 × 1 JJ* of *D. simulans*, were used as the reference strains for comparing DNA sequences. PCR products from the two species were directly sequenced and checked for nucleotide substitution. Any substitutions except those involving G to C are useful for designing ASO probes.
3. DNA target fragments were amplified by PCR and dot-blotted to N^+ nylon filter. The filter was probed with ^{32}P-labeled ASO at appropriate temperatures.
4. Eight individual males from each reference strain were probed to determine whether ASO alleles were fixed in each strain.

At the first stage of the introgression, we used a total of 28 ASO markers, among which 23 had been developed in previous studies (Liu *et al.* 1996; Zeng *et al.* 2000). We made some improvements on 5 of the 23 markers for more sensitive probes. At the second stage, we developed 10 more (Table 1). Unfortunately, two of them, *Hsr* and *kni*, were not fixed in all *D. mauritiana* stocks. A total of 38 ASO markers were used in this study.

Single-fly genotyping: A single fly was ground in 50 μl grinding solution (10 mm Tris pH 8.2, 1 mm EDTA, 25 mm NaCl, 0.2 mg/ml Proteinase K) in a 1.5-ml Eppendorf tube. After a 15-min digestion at room temperature, the tube was incubated at 95° for >2 min and chilled on ice briefly before being stored at −20°.

PCR reactions were set up with 1 μl DNA extraction prepared as above in a volume of 10 μl reaction mixture [2.5 ng/μl primer mix each, 0.25 units of Taq polymerase, 1× Taq buffer, 1.65 mm $MgCl_2$, 150 μm dNTP]. The temperature regimes for all ASO markers were the same except the annealing temperatures (Table 1). For a large number of PCR reactions, 96-well plates were used. The PCR product was denatured with 100 μl alkaline solution (0.4 m NaOH, 15 mm EDTA) and transferred to N^+ nylon filter by using a 96-well dot-blot apparatus. The filter was baked at 80° for >2 hr and then hybridized sequentially with ^{32}P-labeled species-specific ASO probes. After being washed at a temperature appropriate for each ASO probe, signals from the filter were detected by a Molecular Dynamics (Sunnyvale, CA) phosphorimager.

Fertility assay: In the mating test, the number of tester females, duration of mating, and ambient temperature may cause variation in progeny counts (data not shown). To minimize this variation, we used a test regime of 1 male (female) × 3 virgin females (males) for 7 (8) days at room temperature. The number of males tested per line has a mean ± SD of 14 ± 12 with a range of 1–111.

For females, fertility was tested by crossing a single virgin female to three *D. simulans w*; *e* males (<3 days old). On the eighth day, the flies were cleared. Females still alive were collected for genotyping. Vials in which the female and/or all three males were dead were excluded from data analysis. The progeny were sorted and counted by eye color and sex on the thirteenth, sixteenth, and nineteenth days. The number of females tested per line has a mean ± SD of 7 ± 3 with a range of 1–16. A female was considered fertile if she produced any progeny; otherwise she was considered sterile.

Males (<3 days old) were singly crossed with three virgin females of *D. simulans w*; *e* (<3 days old) for 7 days. Females were discarded and males were collected on the seventh day for later genotyping. Offspring were counted by sex and eye color on the thirteenth, sixteenth, and nineteenth days. We included for data analysis only the vials in which the male and at least one female were still alive at the time of collection and the male was the correct genotype. Because the carrying capacity for 7 ml medium in each vial is ~100 (rarely >200) flies, this method can be regarded as quantitative for those males with fertility well below 100 flies.

Male fertility classification: In previous analyses of *X*-chromosome introgressions, fertility was classified primarily on the basis of sperm motility. Davis and Wu (1996) defined a sperm index with four classes based on the number of motile sperm present in the seminal vesicles. Classes 1–4 have 0, 1–25, 25–100, and >100 motile sperm, respectively. Introgression lines were classified as sterile, quasi-sterile, subfertile, and fertile, according to sperm index and qualitative results of multiple-male mating tests.

In our study, fertility was estimated primarily from the number of progeny produced in single-male mating tests, because sperm motility is very difficult to measure quantitatively. However, a calibration experiment was conducted to define fertility classes (by progeny number) that are comparable to the classes

TABLE 1

Summary of ASO markers

ASO marker	Cytological position	PCR condition[a]	PCR primers	ASO probes[b]	ASO wash condition[b,c]
Pdp1	66A	60°/55°	CCTGTCGGAGAACAACATACC	S: GTATAGCGGTTCTTG	37°
			AGGAGGCGAACGAAAATG	M: CAATAGGTATTGGTA	
tra[d]	73A	56°	GAACAAGCGAGAGGGATAGC	S: GTGGATTATACCGAT	50°
			CTTTGGCGGTGGATTATACC	M: ATCGGTACAATCCAC	
Dbp73D	73D	55°	TTAGTCGTCCTTCCTGTGGCAGAG	S: GATCTTGTGACAATG	47°
			ATCAGCACCTTCTTCCGTTGCGTC	M: GATCTTGAGACAATG	
ash1	76B6-9	55°	TTTCATCAGCAGTGGCGGTG	S: TGAATTGTCTGACCA	55°(S)/52°(M)
			CGGCGTTTTGTAATGTCCCTG	M: TGAATTGCCTGACCA	
rdgC[d]	77B	63°	CAAAGACATCGACTCAGCTACG	S: CTCCACCACAGATAA	50°
			CGAACTCTCCACGATGCC	M: TTATCTGCGGAGGAA	
kni	77E2	55°	ACCAGACATGCAAAGTGTGCGG	S: TGTATCTAAAACCTA	45°
			TTCGACATGCCCACGTTGTAG	M: TAGGTTTCAGATACA	
Pka	77F1-5	55°	AAAACCGATGTGAAGCTGTGGGG	S: TGCTAAGACAAATAG	40°
			CGTGATGTTGCGTTTGAGTATGTC	M: TGCAATGACAAATGG	
Edg78	78E	53°	CGAGCACTGTTGTTATTGTTGCC	S: GGATGAGACCCTTGT	57°(S)/52°(M)
			TTCTGGAAGCTGCGGATCTG	M: ACAAGGGACTCATCC	
Rga	83B	55°	GATGGTGAGTGCTGGAACACTTC	S: CAACCTTCCTGTTAC	45°(S)/37°(M)
			GTTAATCAGATATTCGGGTGGCAC	M: CAACTTTCCTGGTAC	
lbl	93E	55°	CTTTCCGCTGCCGATATAGTTG	S: CCTGTTATAAGGAGAA	32°
			ACGTCCTTCTTCAGCTCCTCCATG	M: CCTGTTAAAGGAAAA	
Hsr	93D	55°	GAGGCAGTTATCCAGGATGTAAGG	S: TTCATCGTGTTGCGT	47°(S)/42°(M)
			ATGATCTCAGTCACGGTCGCAGAC	M: TTCATCAATGTTGCGT	
Fsh	90C	55°	TGGCTGCGTATTTCCGTCTG	S: AGAATAGATGAATCG	50°(S)/47°(M)
			CAACTTGATGCGATGGTTCAGG	M: GCATTCAACTATTCT	
Fas1[d]	89E1-2	60°/55°	GCGACTCTTTGAGTATCCGC	S: TACCAATTTTCTATTG	37°
			GGCCAAGATGATGCAAATG	M: CAATAGGTATTGGTA	
Akt1	88B9	55°	GTAATAGCAGCGGCGTTAAGAAAG	S: CTTGTTAATAGTATT	40°(M)/37°(S)
			CGGGAATTTTACCTCCTCCACC	M: CTTGTTACTAGTATT	
Su(Hw)	88B3	55°	CAAGAACTCCCTCCAACAAGCAG	S: TTTCCACACACGTGC	50°(S)/45°(M)
			CGCAAACGATACCCACTGTCATAG	M: TTGCCACACACATGC	
ems	88A1-2	55°	GCCCAGTTTATGCCCAATCC	S: AAATTTTTGTCACTG	42°(S)/37°(M)
			CGATTCGCCAGGATAAGGAAC	M: AAATTTGAGTCTCAT	
Mst[d]	87F	50°	TCCTTTGCCTCTTCAGTCCG	S: AGAGAAATGAAAATG	37°
			TCCACAGGCATAGCATGGTC	M: CATTTCCATTTCTGT	
Odh[d]	86D	55°	CAGTGGTGTTGGGACATGAG	S: GTGGGTTCGTGGAAC	50°
			TGGCCAGCTCAAACTTGTC	M: GTTCCACAAACCCAC	
Crc	86BC	55°	AAAGTGGTGCGGGATCGTTCAC	S: TTACTATAGCTACTC	37°
			CATCCTGAGAGGTCTGAATGCC	M: GAGTTGCTATCGTAA	
tld	96A	55°	GCCGAAGGAGCCAAAGGATG	S: CAACTATTGAAGAGG	45°
			AAACCCGATTGTGTGACCCAG	M: CCTCTTCCATAGTTG	

In addition to 20 ASO markers listed here, 18 more had been developed before (Liu *et al.* 1996; Zeng *et al.* 2000). Their locations are 61A(*Lsp1*), 62A(*ve*), 64B–C(*Acr64B*), 65E(*Dbi*), 66D(*h*), 68E(*CycA*), 70D(*fz*), 71CD(*Eip71CD*), 82C(*5-Ht2*), 84B (*Antp*), 92B(*ninaE*), 85E(*Tub85E*), 85A3-B1(*hb*), 95D(*Rox8*), 97AB(*Ald*), 98B(*Mlc1*), 99D(*jan*), and 100E(*Efld2*).

[a] All PCR reactions followed this format: 94° for 3 min/1 cycle; (94° for 45 sec, annealing temperature for 30 sec, 72° for 15 sec)/40 cycles. If two annealing temperatures are given, the first was used for 10 cycles and the second for 30 cycles.

[b] S, *D. simulans*-specific; M, *D. mauritiana*-specific.

[c] ^{32}P-labeled probes were washed at designated temperatures for 15 min.

[d] Also used in Liu *et al.* (1996) and Zeng *et al.* (2000), but ASO probes were newly developed.

used by Davis and Wu (1996) for the *X* chromosome. In this experiment, both sperm index evaluation and mating tests were performed on males from each of 146 different genotypes. On the seventh day of the mating test, multiple males for each genotype were dissected and the sperm index was scored according to Davis and Wu (1996). Figure 3 shows a plot of the average sperm index *vs.* average progeny number for each genotype. The two measures are well correlated, but some exceptions do occur. For instance, there are genotypes having no detected motile sperm, which nevertheless produced a few offspring and vice versa. However, in this study and in the study of Davis and Wu (1996) no males had many motile sperm but produced no progeny, against the expectation that behavioral problems (*i.e.*, failure to mate) were a common cause of sterility in the mating test.

Here we use a fertility classification system based solely on progeny number that was designed to be comparable to the sperm index system of Davis and Wu (1996): sterile (<0),

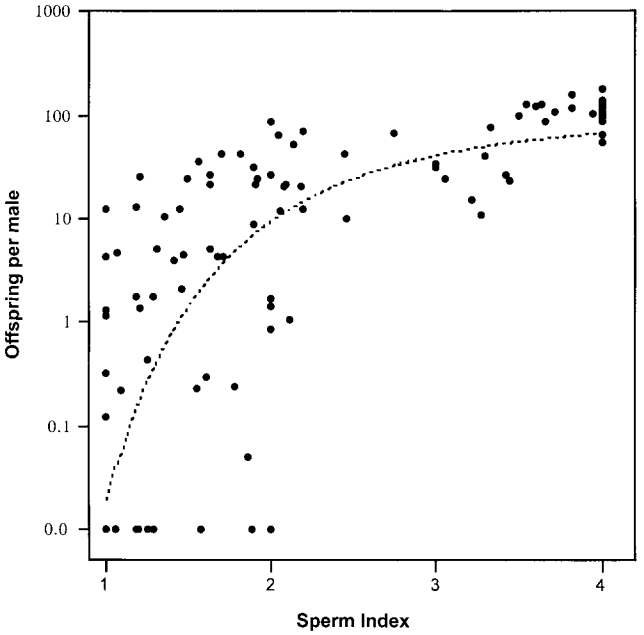

FIGURE 3.—Two ways of measuring male fertility are highly correlated. Individual fertility was assayed consecutively by mating test and sperm motility (sperm index x: 1–4). Progeny counts (y) were transformed in logarithm scale by $\mathrm{Log}_{10}(y + 0.01)$. Each individual of the 146 genotypes was checked for appropriate ASO markers to exclude any recombinants from data analysis. An average of 12.2 and 20.9 individuals per genotype were sampled for scoring sperm index and mating test, respectively. The sample size for mating test was larger because some individuals were not observed for sperm motility. The two measures are highly correlated ($R^2 = 0.62$, $P < 10^{-6}$ after z-transformation; SOKAL and ROHLF 1995, p. 575) and a logistic curve is well fitted ($y = 7.49 - 7.82/(1.41 - 1.53e^{-x})$) ($R^2 = 0.74$, $P \ll 0.001$).

quasi-sterile (0–10), subfertile (10–45), and fertile (>45 progeny per male tested). Table 2 shows that both the distribution of sperm index values within a fertility class and the average for each class are very similar for the autosome (this study) and the *X* chromosome (DAVIS and WU 1996).

Estimating the lengths of introgressed segments: The positions of ASO and $P[w^+]$-insert as landmarks were ordered on the polytene map of *D. melanogaster* and segment lengths were estimated as the number of bands between landmarks (HEINO *et al.* 1994). Each introgressed segment of *D. mauritiana* has up to five landmarks: a central $P[w^+]$-insert flanked by a pair of ASO markers on each side. The left boundary of the segment was generated by recombination between the outer member (now a *D. simulans* allele) and the inner member (still *D. mauritiana*) of the left pair (and similarly for the right boundary). In some cases, the boundary falls between the $P[w^+]$-insert and an ASO marker. For the purpose of length calculation, these boundaries are assumed to lie at the midpoint of the interval defined by each pair of landmarks. Overall, a total of 76 landmarks were used for measuring the introgressions, providing a resolution of ~30 polytene bands per landmark. This level of resolution is comparable to that (~20 bands) on the finely mapped 40% of the *X* chromosome (reviewed in WU and HOLLOCHER 1998).

Statistics: The outcome of a male fertility test for an introgression line is determined by the total cumulative HMS effect, which depends on the number of HMS factors residing in the line and the magnitudes of their individual effects. We would like to estimate these quantities from the observations, namely, the lengths of the introgressed *D. mauritiana* segments and the level of male fertility. More specifically, we wish to estimate the relative densities of HMS factors on the *X vs.* autosomes. For these purposes, we derived a statistical model to account for the HMS accumulation under the following assumptions: The HMS factors have a Poisson distribution along a chromosome; the effect of each factor follows an exponential distribution; and the effects are additive.

TABLE 2

Number of males with various sperm indexes within each fertility class: a comparison between current and previous studies

	Sperm index						
	1	2	3	4	Subtotal	Mean	SD
				DAVIS and WU (1996)			
Sterile	269	15	1	0	285	1.06	0.25
Quasi-sterile	177	67	15	4	263	1.41	0.67
Subfertile	32	55	18	10	115	2.05	0.89
Fertile	6	9	33	102	150	3.54	0.78
				This study			
Sterile	583	19	5	0	607	1.05	0.25
Quasi-sterile	216	150	10	4	380	1.48	0.61
Subfertile	97	185	84	44	410	2.18	0.92
Fertile	7	71	28	279	385	3.50	0.85

The numbers of males with a certain sperm index are compared within each fertility class between previous (DAVIS and WU 1996, Table 4) and current studies. In a Kolmogorov-Smirnov test (SOKAL and ROHLF 1995, p. 438), the D values are 0.0166 for sterile ($P > 0.5$), 0.1046 for quasi-sterile ($P > 0.05$), 0.0687 for subfertile ($P > 0.5$), and 0.1026 for fertile ($P > 0.05$), suggesting the consistence of male fertility classification between these two studies.

In the APPENDIX we have shown that the accumulated effect (z) of HMS on an introgressed segment of length L has a probability density function (pdf),

$$f(z|\theta, \lambda L) = e^{-(\lambda L+\theta z)}\left[\sum_{k=1}^{\infty}\frac{(\lambda L\theta)^k z^{k-1}}{k!(k-1)!}\right] + e^{-\lambda L}\delta(z), \quad (1)$$

where λ and θ are parameters for the Poisson and exponential distributions, respectively. The notation $\delta(z)$ is a function with a value of 1 when $z = 0$ and 0 when $z \neq 0$. The mean and variance for distribution (1) are $\lambda L/\theta$ and $2\lambda L/\theta^2$, respectively. The cumulative density function (cdf) of z on an introgression of length L is

$$F(z|\theta, \lambda L) = \int f(z|\theta, \lambda L)\, dz,$$

which can be obtained numerically.

For estimating parameters in the model, male flies are assigned to one of three classes: fertile when $z < \gamma_1$, semifertile (quasi-sterile and subfertile) when $\gamma_1 < z < \gamma_2$, and sterile when $z > \gamma_2$. The two γ parameters are the thresholds where fertility shifts classes. Note that fertility classes of quasi-sterile and subfertile are pooled because of their small sample sizes. There are two observations for each line: the length of its introgression L_i and its fertility class s_i (where $s_i = S$, SF, or F, indicating whether the line is sterile, semifertile, or fertile, respectively, and $i = 1, 2, \ldots, M$, where M is the total number

of introgression lines tested). We assume that the magnitudes of effect of HMS factors on the *X* have the same distribution as those on the third (*i.e.*, they share the same θ but have different λ's). The likelihood for all observations $O = (s_1, L_1, s_2, L_2 \ldots s_M, L_M)$ can be written as

$$L(O|\theta, \lambda_x, \lambda_3, \gamma_1, \gamma_2) = \prod_{i=1}^{M_X} \{[F(\gamma_1|\lambda_X L_i, \theta)]^F [F(\gamma_2|\lambda_X L_i, \theta) - F(\gamma_1|\lambda_X L_i, \theta)]^{SF} \cdot [1 - F(\gamma_2|\lambda_X L_i, \theta)]^S\} \cdot \prod_{i=1}^{M_3} \{[F(\gamma_1|\lambda_3 L_i, \theta)]^F [F(\gamma_2|\lambda_3 L_i, \theta) - F(\gamma_1|\lambda_3 L_i, \theta)]^{SF} \cdot [1 - F(\gamma_2|\lambda_3 L_i, \theta)]^S\}, \quad (2)$$

where $M = M_X + M_3$. M_3 and M_X are the numbers of introgression lines for the third and the *X* chromosomes, respectively, while λ_3 and λ_X are the corresponding parameters of the Poisson distribution. For each line, 1 or 0 is assigned to each of the three logical variables *S*, *SF*, and *F* according to the fertility observed.

The five parameters in the likelihood function are not identifiable. However, if one of the γ parameters (say γ_1) is fixed as a scalar, the remaining four parameters can be uniquely determined (identifiable). Although these four parameter values depend on the scalar chosen, the biologically important ratios λ_X/θ, λ_3/θ, and λ_X/λ_3 are independent of the scalar. Equation 2 obviously can be modified to fit the observations on the *X* and the third chromosomes separately, as

$$L(O|\theta, \lambda_j, \gamma_1, \gamma_2) = \prod_{i=1}^{M_j} \{[F(\gamma_1|\lambda_j L_i, \theta)]^F [F(\gamma_2|\lambda_j L_i, \theta) - F(\gamma_1|\lambda_j L_i, \theta)]^{SF} \cdot [1 - F(\gamma_2|\lambda_j L_i, \theta)]^S\}, \quad (3)$$

where *j* indexes the *X* or third chromosomes.

We maximized the likelihood (Equations 2 or 3) by using the observations from the *X* and third chromosomes either jointly or separately through a Markov chain Monte Carlo (MCMC) technique, using the Metropolis algorithm (Gelman *et al.* 1995, chap. 11). We arbitrarily set $\gamma_1 = 1$ and chose initial values for the other four parameters (θ, λ_X, λ_3, and γ_2). The parameter estimates that maximize the likelihood functions were obtained from the MCMC after 1000 iterations of burn-ins. In the MCMC simulation, the absolute values of θ, λ_X, and λ_3 sometimes did not converge well, but the value of γ_2 and the ratios of λ_X/θ and λ_3/θ always converged nicely.

RESULTS

Spontaneous mutations distinguished from hybrid incompatibilities: Sterility and lethality detected by the protocol of Figure 2A could be due to spontaneous mutations rather than to hybrid incompatibility. The test protocol in Figure 2B was used to distinguish these two situations. If spontaneous mutations arose in independent lines during the course of the experiment, they were unlikely to be allelic to each other. Several lines were found to contain spontaneous mutations: two homozygous male sterile lines (31.1 and 31.2), two homozygous female sterile lines (51.4, 51.11), and seven homozygous lethal lines (16.3, 20.2, 20.7, 20.11, 20.16, 20.23, and 20.25). Since most of the spontaneous mutations were clustered within a particular $P[w^+]$-insert line (as indicated by the number preceding the dot in the line name), there may be only four independent mutations (one male sterile in line 31, one female sterile in line 51, and one lethal in each of lines 16 and 20). Thus, for the 38 $P[w^+]$-insert lines used in this study, the spontaneous mutation rate is on the order of 2.5% for male or female sterility and is on the order of 5% for lethality.

"Faster male" evolution—much larger accumulation of HMS than of hybrid female sterility: In this study, we tested 218 introgression lines for homozygous male fertility and 173 lines for homozygous female fertility. Among the 145 lines tested for fertility in both sexes, 67 lines are male infertile (30 steriles, 22 subfertiles, 15 quasi-steriles) while none are female infertile. Thus, there is a clear difference in density of male *vs.* female hybrid sterility factors. A similar difference in fertility between the sexes was reported previously by True *et al.* (1996b). However, in the earlier experiment, there was a possible bias because the lines were established by repeated backcrossing of heterozygous females. This protocol may have selected against factors with a heterozygous effect on female fertility. Furthermore, the extent of coverage of the autosomes by introgressions was not determined, so the possible magnitude of selection was unknown. In the experiment reported here, the lengths of introgressions were defined by ASO and $P[w^+]$ landmarks and it was determined that these segments cover the entire third chromosome, eliminating any possible bias in coverage. Therefore, the density difference in male *vs.* female sterility is now very well established.

The density of HMS factors on the *X* is estimated to be ~2.5 times higher than that on the third chromosome: In this study we also observed the lengths of the third chromosome introgressions tested for male and female fertility as homozygotes. In Figure 4, we graphically show the 218 lines tested for male fertility. Similar observations on 265 introgressions, which cover ~60% of the *X* chromosome, were obtained from previous studies (Perez *et al.* 1993, Figure 4; Cabot *et al.* 1994, Figure 5; Perez and Wu 1995, Figure 2; Davis and Wu 1996, Figure 3). The descriptive statistics are summarized in Table 3. The *X* chromosome appears to have a substantially higher number and/or larger effects of individual HMS factors than the third chromosome. Only very short introgressions on the *X* can be fertile and lengths that are sterile on the *X* frequently are fertile on the third chromosome.

The phenotypic effect on fertility of an introgressed segment depends on two variables: the number of factors and the magnitudes of their effect. Assuming that the number of factors has a Poisson distribution with parameter λ and that the magnitude of effect has an exponential distribution with parameter θ, we have derived a distribution of the cumulative effect that incorporates both variables (appendix) and estimated the parameters with Markov chain Monte Carlo simulations. Since θ and λ are confounded in the observations, we assume that θ is the same for both the *X* and third chromosomes and estimate the relative densities of factors on

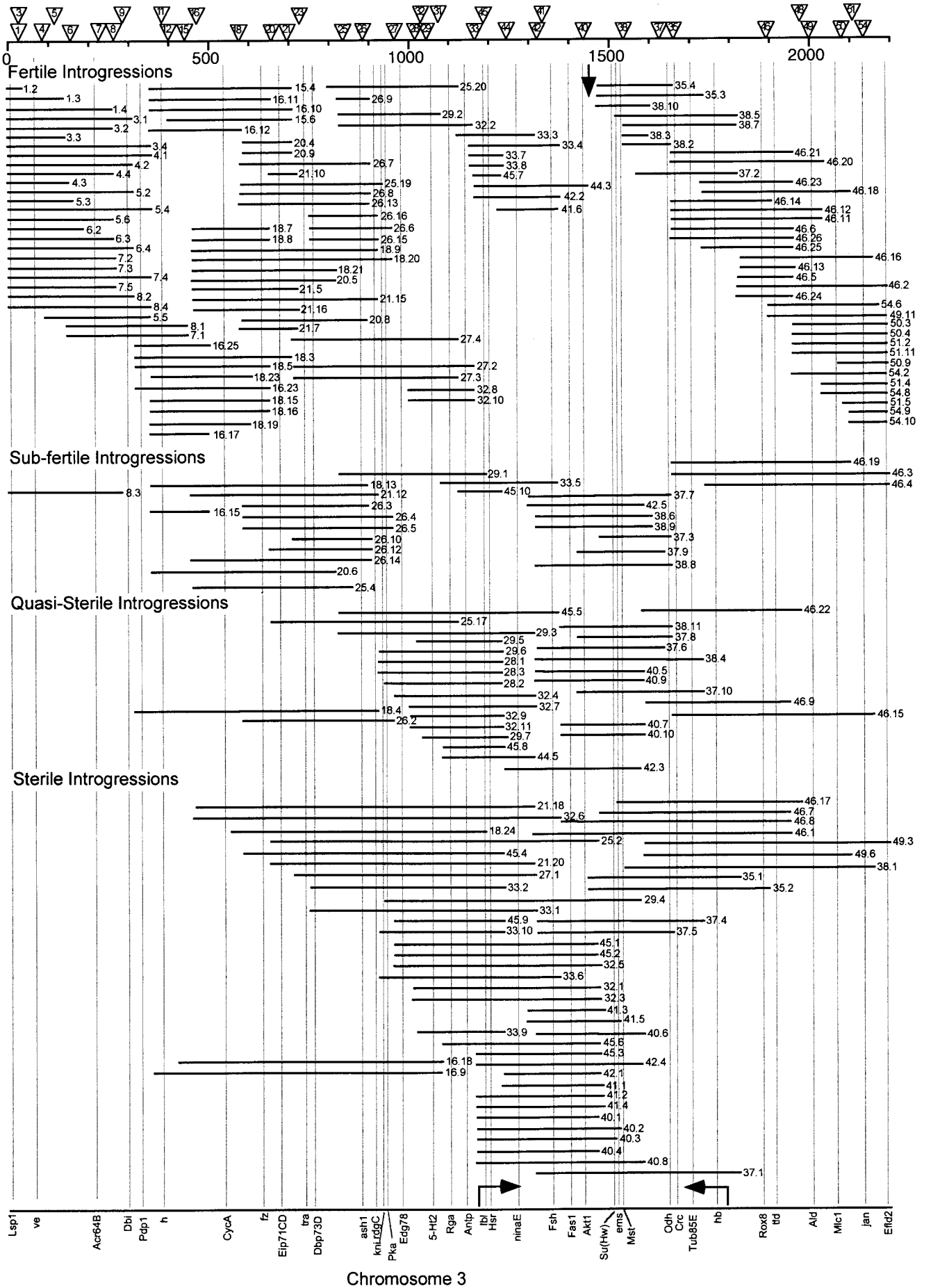
0
500
1000
1500
2000
Fertile Introgressions
Sub-fertile Introgressions
Quasi-Sterile Introgressions
Sterile Introgressions
Lsp1
ve
Acr64B
Dbi
Pdp1
h
CycA
fz
Eip71CD
tra
Dbp73D
ash1
kni
rdgC
Pka
Edg78
5-Ht2
Rga
Antp
lbl
Hsr
ninaE
Fsh
Fas1
Akt1
Su(Hw)
ems
Mst
Odh
Crc
Tub85E
hb
Rox8
ttd
Ald
Mlc1
jan
Efid2
Chromosome 3

TABLE 3

Profiles of introgression lengths on the third and *X* chromosomes

	Male fertility				Female fertility:
	Fertile	Subfertile	Quasi-sterile	Sterile	Fertile
			Chromosome III		
No. lines	115	25	30	48	173
Mean ± SD	251 ± 104	335 ± 118	324 ± 117	462 ± 183	334 ± 155
Range	44–501	115–540	158–540	138–924	44−924
			Chromosome *X*		
No. lines	209	24	3	29	—
Mean ± SD	53 ± 41	38 ± 12	200 ± 0	163 ± 75	—
Range	11–200	24–70	200	70–260	—

Introgression lengths are in numbers of polytene bands (mean ± SD). The introgression lengths on the two chromosomes are very different for all four classes (*t*-test, $P \ll 0.001$ for all classes). Data of the *X* chromosome are compiled from Perez *et al.* (1993), Figure 4; Cabot *et al.* (1994), Figure 5; Perez and Wu (1995), Figure 2; Davis and Wu (1996), Figure 3.

these two chromosomes (λ_X/λ_3), as well as some other informative ratios of parameters.

Using quite different initial values, the MCMC simulation was run three times, each with 2000 iterations, either jointly or separately for the *X* and the third chromosomes. The estimates of γ_2, λ_X/θ, λ_3/θ, and λ_X/λ_3 converged very well (Table 4). The average ratio of λ_X/λ_3 is estimated as ~2.5, which indicates that the *X* chromosome has ~2.5 times the density of HMS factors as the third chromosome.

The number of HMS equivalents in the genome is ~15: An HMS "equivalent" is the minimum value of *z* (the cumulative negative effect on fertility) required to cause complete sterility (*i.e.*, γ_2 in our model). The number of HMS equivalents on a chromosome can be calculated as the expected value of *z* for the entire length of the chromosome (*i.e.*, $\lambda L/\theta$, where *L* is the number of bands in the chromosome) divided by γ_2. Table 4 shows that the *X* and third chromosomes each have about five HMS equivalents. Although the *X* has a twofold higher density of HMS factors, it is about one-half the total length of the third chromosome ($L_3 = 2200$ and $L_X = 981$ in polytene bands; Heino *et al.* 1994). Therefore, the *X* has roughly the same accumulation of HMS effects as the third chromosome (under our model where the distribution of individual factor effects does not differ between chromosomes). Since the second chromosome is about the same length as the third and the fourth chromosome is tiny, the total number of HMS equivalents in the genome is ~15. Similarly, the number of hybrid male *semi*sterile equivalents per chromosome is $\lambda L/\theta\gamma_1$ and the total number for the genome is ~25.

Each HMS equivalent could be considered as an introgressed *D. mauritiana* segment, nonoverlapping with other such segments and consisting of a set of "minor" HMS factors that together cause complete sterility. In a companion article (Tao *et al.* 2003, this issue), we show by quantitative trait loci (QTL) mapping that the third chromosome contains 19 minor sterility factors. Similarly, 9 minor factors have been mapped on ~40% of the *X* chromosome (Wu and Hollocher 1998). Thus the total number of the QTL (or minor genes) in the genome is ~60 and, on average, every HMS equivalent

Figure 4.—Homozygous male fertility of introgressions on the third chromosome. Bars represent the introgressed *D. mauritiana* segments in an otherwise pure *D. simulans* background. The boundaries for each introgressed segment are defined by ASO markers as well as $P[w^+]$-inserts. Serial numbers of sublines are labeled at the right of bars, with the digits before dots corresponding to $P[w^+]$-inserts (open triangles at top). A total of 218 lines were tested, including 25 lines that had already been made and kept as homozygotes. The ASO markers and $P[w^+]$-inserts are positioned according to polytene bands. Fertile introgressions cover the whole third chromosome except one gap (solid arrow). Males with homozygous introgressions covering this gap had their fertility severely reduced and sired predominantly female offspring (Tao *et al.* 2001). A total of 38 *D. mauritiana* $P[w^+]$-insert lines were used in this study. These lines have been described before (True *et al.* 1996a, Table A1): 1 [61CD(I)], 3 [61CD(II)], 4 [62BC], 5 [62E], 6 [63C], 7 [64C], 8 [65A], 9 [65CD], 11 [66D], 12 [66EF], 15 [67BC(II)], 16 [67DE], 18 [69D], 20 [71A(II)], 21 [71F72A], 23 [72EF(II)], 25 [75C(II)], 26 [76C], 27 [78CD], 28 [79F], 32 [83AB], 29 [82A], 31 [82E], 33 [84DE], 45 [93EF], 44 [92E], 42 [91BC], 41 [91A], 40 [90B], 38 [87EF], 37 [86EF], 35 [86AB], 46 [95A(I)], 48 [96E], 49 [97A], 50 [98DE], 51 [99A(I)], and 54 [99D]. Note that the cytological positions are based on the *D. melanogaster* map. There is an inversion at 84F–93F between *D. melanogaster* and the *D. simulans* triad (Lemeunier and Ashburner 1976; indicated by bent arrows facing each other). The cytological locations of 32 and 40 determined by True *et al.* (1996a) were not correct. The current locations of 32 and 40 were inferred from meiotic recombination data obtained in this study: 32 is between 28 and 29, and 40 is distal to ASO marker *Akt1* (89B9).

TABLE 4

The MCMC output for three runs

No. runs[a]	Methods[b]	λ_X/θ	λ_3/θ	γ_2	λ_X/λ_3	$S_X(\frac{1}{2})$[c]	$S_X(1)$[c]	$S_3(\frac{1}{2})$[c]	$S_3(1)$[c]
1	*X*	0.0089 ± 0.0009	—	1.7 ± 0.2	—	8.8 ± 0.9	5.2 ± 0.6	—	—
	3	—	0.0036 ± 0.0002	1.7 ± 0.1	—	—	—	7.9 ± 0.5	4.6 ± 0.3
	X and 3	0.0094 ± 0.0007	0.0038 ± 0.0003	1.8 ± 0.2	2.5 ± 0.3	9.2 ± 0.7	5.1 ± 0.6	8.4 ± 0.7	4.6 ± 0.3
2	*X*	0.0094 ± 0.001	—	1.8 ± 0.2	—	9.3 ± 1.0	5.2 ± 0.6	—	—
	3	—	0.0035 ± 0.0002	1.7 ± 0.1	—	—	—	7.8 ± 0.4	4.7 ± 0.3
	X and 3	0.0092 ± 0.0007	0.0038 ± 0.0002	1.8 ± 0.1	2.4 ± 0.2	9.0 ± 0.7	5.1 ± 0.4	8.3 ± 0.5	4.7 ± 0.3
3	*X*	0.0093 ± 0.0009	—	1.7 ± 0.2	—	9.1 ± 0.9	5.4 ± 0.6	—	—
	3	—	0.0035 ± 0.0002	1.6 ± 0.1	—			7.7 ± 0.4	4.7 ± 0.3
	X and 3	0.0092 ± 0.0007	0.0037 ± 0.0003	1.8 ± 0.1	2.5 ± 0.3	9.1 ± 0.7	5.2 ± 0.5	8.2 ± 0.6	4.7 ± 0.3

All estimates (mean ± SD) from the MCMC methods are summarized from the second half of 2000 iterations.

[a] Initial values for MCMC (±SD): run 1, $\theta = 3.0$ (±0.1), $\lambda_X = 0.03$ (±0.001), $\lambda_3 = 0.01$ (±0.0005), $\gamma_2 = 3.0$ (±0.1); run 2, $\theta = 1.5$ (±0.15), $\lambda_X = 0.035$ (±0.0015), $\lambda_3 = 0.005$ (±0.0015), $\gamma_2 = 3.5$ (±0.15); run 3, $\theta = 1.0$ (±0.1), $\lambda_X = 0.01$ (±0.001), $\lambda_3 = 0.035$ (±0.001), $\gamma_2 = 3.5$ (±0.1).

[b] The MCMC simulations are run separately for observations from the *X* (*X*) or the third (3) chromosomes or jointly (*X* and 3).

[c] $S_X(\frac{1}{2})$, the estimated number of hybrid male *semi*sterility equivalents on the *X* chromosome, calculated as $\lambda_X L_X/\theta$, where L_X (= 981) is the polytene band count on the *X* chromosome. $S_X(1)$, the estimated number of hybrid male *full*-sterility equivalents on the *X* chromosome, calculated as $\lambda_X L_X/\theta/\gamma_2$. $S_3(\frac{1}{2})$ and $S_3(1)$ are similarly calculated with $L_3 = 2200$.

consists of ~4 minor genes (*i.e.*, 60 minor genes/15 HMS equivalents). In other words, ~4 minor genes together can cause complete sterility. By the same argument, one semisterile equivalent consists of ~2.5 minor genes, which is consistent with the observation that adding one or two minor genes to a semifertile genotype can make it completely sterile (Tao *et al.* 2003).

DISCUSSION

More hybrid male sterility than female sterility and lethality: The much higher prevalence of hybrid male compared to female sterility and inviability has been reported previously (Hollocher and Wu 1996; True *et al.* 1996b). Here we support the earlier conclusions with some technical improvements in the experimental design. A potential bias against discovering female sterility in the study of True *et al.* (1996b) was removed by defining the lengths of introgressions and covering the entire third chromosome. Also, spontaneous mutations that occur during the experiment were identified and their frequencies account approximately for the low frequencies of female sterility and lethality observed by True *et al.* (1996b; 7 and 5%, respectively). Thus, there is no clear evidence that a single introgression causes female sterility or lethality due to hybrid incompatibility in *D. simulans* and *D. mauritiana*. However, there is evidence for an accumulation of minor female sterility factors. In a study by Davis *et al.* (1994), homozygous introgressions on either the second or the third chromosome were female fertile, but simultaneous homozygosity on both chromosomes resulted in female sterility. Several possible causes of the difference in male *vs.* female hybrid sterility in Drosophila have been discussed at length in the literature, such as sexual selection for male reproductive characters and special properties of spermatogenesis (Wu and Davis 1993).

Reliability of the large *X* effect: A major conclusion of this study is that the *X* chromosome appears to contain ~2.5 times the density of hybrid male sterility factors as the autosomes. Three caveats regarding this conclusion must be considered.

1. The observations on the *X*-linked and autosomal HMS were made in two different laboratories using different methodologies. The classification of the *X*-linked introgressions was based largely on sperm motility whereas that for the third chromosome was based on progeny number in mating tests. However, the progeny number classification system used here was designed to give equivalent results for the two systems and the calibration experiment reported here gives no indication of a bias toward detecting more sterility on the *X*.

   Furthermore, this caveat concerning different methodologies has been addressed to some extent in the earlier study of True *et al.* (1996b). In that study, *P*-element-marked introgressions were made throughout the *X* and both major autosomes. The fraction of hemizygous *X* introgressions that were male sterile was ~50% greater than the fraction of homozygous autosomal segments. Unfortunately, the lengths of introgressed segments were not estimated directly in the True *et al.* (1996b) study, so the density difference could not be estimated quantitatively from those data and a possible difference in introgression length distributions could not be ruled out. Nevertheless, the True *et al.* (1996b) results qualitatively support the conclusions of the study reported here.

2. The ~2.5-fold *X vs.* autosome difference reported here may be an overestimate due to biased sampling of the *X* chromosome. The three *X* chromosomal regions analyzed by Wu and colleagues had previously been implicated as hosting major HMS factors (Coyne and Charlesworth 1986, 1989), which could mean that they contain a higher density than the rest of the *X* chromosome. However, the distribution of sterile introgressions on the *X* in the study of True *et al.* (1996b) does not support this notion (see their Figure 4). Although there is a suggestion of more sterility near the centromere and less at the distal tip, this is probably an artifact of introgression length. Introgressions near the centromere may tend to be longer because of reduced recombination in that region, while those near the tip are shorter because of the edge effect. Consistent with this interpretation, only two HMS loci are finely mapped in one proximal region (Perez *et al.* 1993) while at least one HMS locus is mapped on the tip of the *X* chromosome (Maside *et al.* 1998). A similar pattern is apparent for the third chromosome in Figure 4 of True *et al.* (1996b), even though genetic analyses indicate a fairly even distribution of HMS factors (Tao *et al.* 2003). Therefore, the sampling bias of the *X* chromosome may be very slight at most. We also note again that the study of True *et al.* (1996b) showed that the fraction of sterile *X* introgressions is ~50% greater than the fraction of sterile third chromosome introgressions, even though the two chromosomes were sampled by the same methods in that study. Therefore, an excess of HMS factors on the *X* is supported by both studies, although the quantitative estimate of ~2.5-fold reported here may represent an upper limit.
3. The estimate of relative densities on the *X* chromosome and autosomes depends on an assumption that the distributions of magnitude of effect are the same, which may not be the case. Genes on the *X* and autosomes have different evolutionary dynamics (Charlesworth *et al.* 1987; Orr and Betancourt 2001) that may lead to differences in either substitution rates or, possibly, the average effects of mutations that go to fixation. In either case, however, the observation of a difference in the load of HMS factors suggests that the *X* chromosome may play a special role in the evolution of hybrid incompatibility.

Implications of differential density in HMS factors: There are two basic interpretations of a higher *X vs.* autosome density of HMS factors: (1) the rate of substitution of mutations is higher at *X*-linked than at autosomal loci or (2) the *X* chromosome contains a higher density of genes that have potential effects on male fertility. In humans, sex and reproduction-related genes seem to be disproportionately represented on the *X* chromosome (reviewed in Saifi and Chandra 1999). However, two lines of evidence suggest that this is not the case in Drosophila. First, mutagenesis screens for male sterility have not revealed a biased distribution between the *X* and autosomes (Lindsley and Tokuyasu 1980). Second, there is evidence for a paucity of genes on the *X* that show male-biased expression (Parisi *et al.* 2003), indicating perhaps a deficiency of male-fertility-related genes on the *X* relative to autosomes. Therefore, an excess density of HMS factors on the *X* is most likely due to different substitution rates.

A higher rate of substitution on the *X* implies that evolutionary divergence leading to reproductive isolation occurs through selection rather than through random drift, because the expected substitution rate for neutral alleles is simply the mutation rate for both *X* and autosomal genes. However, it is possible that an *X vs.* autosome bias in mutation rate could result from a sex bias in mutation rate, since the *X* is overrepresented in the homogametic sex. There is evidence for a higher mutation rate in males for mammals (Shimmin *et al.* 1993; Makova and Li 2002) and birds (Ellegren and Fridolfsson 1997), but this would tend to make mutation rates on the *X* lower than those on the autosomes. In Drosophila, no evidence is found to support a sex-biased mutation rate (Bauer and Aquadro 1997). There is also no evidence for a higher rate of substitution on the *X vs.* autosomes from sequence divergence studies (Betancourt *et al.* 2002), consistent with the notion that a majority of substitutions are neutral.

A higher rate of substitution on the *X* also implies that certain types of selection may occur more frequently than others. For example, population genetic theory shows that the substitution rate of newly arisen *X*-linked mutations is higher than that of new autosomal mutations, provided that the mutations are recessive or partially recessive (Charlesworth *et al.* 1987). However, the opposite result is obtained by considering adaptation to a sudden environmental change in which mutations that were previously deleterious and held in mutation-selection balance become beneficial and fixed by selection. In that case, *X*-linked genes evolve somewhat more slowly than autosomal genes, regardless of dominance (Orr and Betancourt 2001). Thus, observation of a higher rate of substitution on the *X* would seem to favor the fixation of new mutations rather than the fixation of standing variation in the evolution of reproductive isolation. Perhaps a change in the direction of selective effect on a mutation is unusual, even though there are some well-documented cases, such as industrial melanism in moths (Kettlewell 1973).

If the estimated *X vs.* autosome ratio of ~2.5:1 for HMS accumulation is equivalent to the relative fixation probability for newly arisen favorable mutations (R_X in Charlesworth *et al.* 1987), we can derive some speculations about the degree of dominance of those mutations. Assuming a simple situation where selection is acting equally on the two sexes and d is the degree of

dominance, R_X is roughly $(2d + 1)/4d$, showing that $R_X > 1$ when $d < 0.5$ (recessive or partially recessive mutations). Similarly, for genes expressed only in the heterogametic sex, as may be the case for a majority of male fertility genes, R_X is $\sim 1/2d$. Thus, a value of $d = 0.2$ gives the $R_X = 2.5$ estimated here. This calculation implies that advantageous mutations fixed for spermatogenesis are moderately recessive.

The invasion of sex-ratio meiotic drive and its subsequent suppression is an additional mechanism that could account for the large *X* effect, as well as the rapid evolution of male (relative to female) hybrid sterility found in Drosophila. When the *X* and *Y* chromosomes in the heterogametic sex are not recombining (as is usually the case), sex-linked segregation distorters can invade the population more readily than autosomal ones can (Frank 1991; Hurst and Pomiankowski 1991). Since sex-ratio distortion is not stable, suppressors (often autosomal) may also invade the population readily, rendering the distortion cryptic. Multiple cycles of distortion and suppression may occur, leading to genetic divergence that is distributed on both the *X* and autosomes, but tending to be more heavily represented on the *X* where the distortion initiates (unless multiple autosomal suppressors are required for each *X*-linked distorter). This genetic divergence may lead to sterility through disturbance of spermatogenesis in hybrids when the cryptic systems may become unsuppressed. The prevalence of this mechanism is not clear yet, but two apparent cases of cryptic sex-ratio drive in *D. simulans* (Dermitzakis *et al.* 2000; Tao *et al.* 2001) and another in *D. pseudoobscura* (Orr and Presgraves 2000) have been reported. In one case, a suppressor is also causing hybrid male sterility (Tao *et al.* 2001). An attractive aspect of this meiotic drive theory is that it predicts a faster evolution of factors affecting the heterogametic sex, consistent with Haldane's rule. Indeed, a recent literature survey appears to show that there is faster female evolution in Lepidoptera, which have heterogametic females (Presgraves 2002).

If selective fixation of recessive mutations is the predominant cause of the "larger *X*" effect on HMS, we might also expect to see such an effect on other kinds of traits. However, if it is due to a more specialized mechanism such as the evolution of meiotic drive systems, the effect may be more restricted to traits related to spermatogenesis. The following summary of studies in Drosophila and Lepidoptera indicates that more work is needed to determine whether a large *X* effect occurs for other traits.

In Drosophila, a larger *X* effect is not clearly evident for traits other than hybrid male sterility (Charlesworth *et al.* 1987). In some cases a larger *X* effect has been suggested for behavior (*e.g.*, Tan 1946; Ewing 1969; Kawanishi and Watanabe 1981; Kyriacou and Hall 1986; reviewed by Reinhold 1998) and for morphology (*e.g.*, Templeton 1977; Val 1977). However, the evidence is not strong because the level of genetic resolution in most studies is low and the number of genes affecting any one trait also may be low. In a relatively high-resolution study of interspecific divergence in genital morphology between *D. simulans* and *D. mauritiana*, the *X* was not overrepresented. About 10% (2/19) of QTL were on the *X* (Zeng *et al.* 2000), whereas the *X* represents ~20% of euchromatic DNA. The morphological trait analyzed in that study (size and shape of the genital arch) represents the only qualitatively distinct morphological difference between those species.

There is stronger evidence for a large *X* effect on behavioral and morphological traits in Lepidoptera (reviewed by Sperling 1994; Prowell 1998; also see recent cases in Jiggins *et al.* 2001; Naisbit *et al.* 2002). Unlike Drosophila where the *X* chromosome is large, the *X* chromosome in Lepidoptera is generally very small, accounting for only ~3% of the genome. Therefore, it appears unlikely, just by chance, that out "of a total of 77 traits, 39% are *X*-linked, 10% are maternal or *Y*-linked, and 51% are autosomally inherited" (Prowell 1998, p. 311). Many of these studies confound dominance with hemi/homozygous effect in comparing the *X* and autosomes, but even so, the degree of dominance would have to be considerably <0.20 to account for the large overrepresentation of the *X*. Unfortunately, there are no detailed genetic dissections of hybrid sterility in Lepidoptera, so we do not know whether a large *X* effect on that trait prevails in species where the female is heterogametic.

The results of the study reported here support the notion that *X*-linked genes play a special role in the evolution of reproductive isolation in Drosophila. Two possible mechanisms are the evolution of sex-ratio distortion systems and selective fixation of advantageous recessive mutations. Additional work on the generality of the large *X* effect across different traits and groups of organisms is likely to provide further understanding of these mechanisms.

We thank Lynn Stam, Kristen Inoue, Jermaine Fuller, Lixia Jiang, and Jing Zhang for technical support during this study; J. A. Coyne for stocks; A. W. Davis and C.-I Wu for communicating on measuring male fertility; and D. A. Barbash, N. A. Johnson, C. H. Langley, S. V. Nuzhdin, M. Turelli and the UC Davis Evolution Discussion Group, J. Blumenstiel, R. J. Kulathinal, C. Castillo-Davis, C. Meiklejohn, J. Parsch, and J. Ranz for critical reading of various drafts of this article. We are grateful to two anonymous reviewers whose suggestions have improved this article in various aspects. We also thank Cecelia Laurie for advice on the MCMC technique. Y.T. thanks R. Fehon, D. Kiehart, J. Mercer, M. Rausher, and M. Uyenoyama for guidance. Y.T. was a recipient of a Catherine Stern Dissertation Year Fellowship of Duke University. This work was supported by National Institutes of Health grants 47292 and 65169.

LITERATURE CITED

Bateman, A. J., 1959 The viability of near-normal irradiated chromosomes. Int. J. Radiat. Biol. **1:** 170–180.

Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **14:** 1252–1257.

Betancourt, A. J., D. C. Presgraves and W. J. Swanson, 2002 A

test for faster *X* evolution in Drosophila. Mol. Biol. Evol. **19:** 1816–1819.

Cabot, E. L., A. W. Davis, N. A. Johnson and C.-I Wu, 1994 Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics **137:** 175–189.

Charlesworth, B., J. A. Coyne and N. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Coyne, J. A., and B. Charlesworth, 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 B. Charlesworth, 1989 Genetic analysis of *X*-linked sterility in hybrids between three sibling species of Drosophila. Heredity **62:** 97–106.

Coyne, J. A., and H. A. Orr, 1989 Two rules of speciation, pp. 180–207 in *Speciation and Its Consequences*, edited by D. Otte and J. A. Endler. Sinauer Associates, Sunderland, MA.

David, J., C. Bocquet, F. Lemeunier and L. Tsacas, 1976 Persistence of male sterility in strains issued from hybrids between two sibling species: *Drosophila simulans* and *D. mauritiana*. J. Genet. **62:** 93–100.

Davis, A. W., and C.-I Wu, 1996 The broom of the sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of Drosophila. Genetics **143:** 1287–1298.

Davis, A. W., E. G. Noonburg and C.-I Wu, 1994 Evidence for complex genic interactions between conspecific chromosomes underlying hybrid female sterility in the *Drosophila simulans* clade. Genetics **137:** 191–199.

Dermitzakis, E. T., J. P. Masly, H. M. Waldrip and A. G. Clark, 2000 Non-Mendelian segregation of sex chromosomes in heterospecific Drosophila males. Genetics **154:** 687–694.

Dobzhansky, Th., 1936 Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21:** 113–135.

Ellegren, H., and A.-K. Fridolfsson, 1997 Male-driven evolution of DNA sequences in birds. Nat. Genet. **17:** 182–184.

Ewing, A. E., 1969 The genetic basis of sound production in *Drosophila pseudoobscura* and *D. persimilis*. Anim. Behav. **17:** 555–568.

Frank, S. A., 1991 Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution **45:** 262–267.

Fry, J. D., P. D. Keightley, S. L. Heinsohn and S. V. Nuzhdin, 1999 New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **96:** 574–579.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 1995 *Bayesian Data Analysis*. Chapman & Hall, London.

Haldane, J. B. S., 1924 A mathematical theory of natural and artificial selection. Part I. Trans. Camb. Philos. Soc. **23:** 19–41.

Heino, T. I., A. O. Saura and V. Sorsa, 1994 Maps of the salivary gland chromosomes of *Drosophila melanogaster*. Dros. Inf. Serv. **73:** 621–738.

Hey, J., and R. M. Kliman, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

Hollocher, H., and C.-I Wu, 1996 The genetics of reproductive isolation in the *Drosophila simulans* clade: X *vs.* autosomal effects and male *vs.* female effects. Genetics **143:** 1243–1255.

Hurst, L. D., and A. Pomiankowski, 1991 Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics **128:** 841–858.

Jiggins, C. D., M. Linares, R. E. Naisbit, C. Salazar, Z. H. Yang *et al.*, 2001 Sex-linked hybrid sterility in a butterfly. Evolution **55:** 1631–1638.

Kawanishi, M., and T. K. Watanabe, 1981 Genes affecting courtship song and mating preference in *Drosophila melanogaster*, *Drosophila simulans* and their hybrids. Evolution **35:** 1128–1133.

Keightley, P. D., 1996 Nature of deleterious mutation load in Drosophila. Genetics **144:** 1993–1999.

Keightley, P. D., 1998 Inference of genome-wide mutation rates and distribution of mutation effects for fitness traits: a simulation study. Genetics **150:** 1283–1293.

Keightley, P. D., and O. Ohnishi, 1998 EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. Genetics **148:** 753–766.

Kettlewell, B., 1973 *The Evolution of Melanism: The Study of a Recurring Necessity*. Clarendon Press, Oxford.

Kyriacou, C. P., and J. C. Hall, 1986 Interspecific genetic control of courtship song production and reception in *Drosophila*. Science **232:** 494–497.

Lemeunier, F., and M. Ashburner, 1976 Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc. R. Soc. Lond. Ser. B **193:** 275–294.

Lindsley, D. L., and K. T. Tokuyasu, 1980 Spermatogenesis, pp. 226–294 in *The Genetics and Biology of Drosophila*, Vol. 2d, edited by M. Ashburner and T. R. F. Wright. Academic Press, New York.

Lindsley, D. L., and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

Liu, J. J., J. M. Mercer, L. F. Stam, G. C. Gibson, Z-B. Zeng *et al.*, 1996 Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. Genetics **142:** 1129–1145.

Makova, K. D., and W-H. Li, 2002 Strong male-driven evolution of DNA sequences in human and apes. Nature **416:** 624–626.

Maside, X. R., J. P. Barral and H. F. Naveira, 1998 Hidden effects of *X* chromosome introgressions on spermatogenesis in *Drosophila simulans* × *D. mauritiana* hybrids unveiled by interactions among minor genetic factors. Genetics **150:** 745–754.

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

Naisbit, R. E., C. D. Jiggins, M. Linares, C. Salazar and J. Mallet, 2002 Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. Genetics **161:** 1517–1526.

Orr, A. H., and A. J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics **157:** 875–884.

Orr, A. H., and D. C. Presgraves, 2000 Speciation by postzygotic isolation: forces, genes and molecules. BioEssays **22:** 1085–1094.

Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley *et al.*, 2003 Paucity of genes on the *Drosophila X* chromosome showing male-biased expression. Science **299:** 697–700.

Perez, D. E., and C.-I Wu, 1995 Further characterizations of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. Genetics **140:** 201–206.

Perez, D. E., C.-I Wu, N. A. Johnson and M.-L. Wu, 1993 Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (*Ods*). Genetics **134:** 261–275.

Presgraves, D. C., 2002 Patterns of postzygotic isolation in Lepidoptera. Evolution **56:** 1168–1183.

Prowell, D. P., 1998 Sex linkage and speciation in Lepidoptera, pp. 309–319 in *Endless Forms: Species and Speciation*, edited by D. J. Howard and S. H. Berlocher. Oxford University Press, New York.

Reinhold, K., 1998 Sex linkage among genes controlling sexually selected traits. Behav. Ecol. Sociobiol. **44:** 1–7.

Rice, W. R., 1984 Sex chromosome and the evolution of sexual dimorphism. Evolution **38:** 735–742.

Saifi, G. M., and H. S. Chandra, 1999 An apparent excess of sex- and reproduction-related genes on the human *X* chromosome. Proc. R. Soc. Lond. Ser. B **266:** 203–209.

Saiki, R. K., T. L. Bugawan, G. T. Horn, K. B. Mullis and H. A. Erlich, 1986 Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature **324:** 163–166.

Searle, S. R., 1971 *Linear Models*. Wiley, New York.

Shimmin, L. C., B. H.-J. Chang and W.-H. Li, 1993 Male-driven evolution of DNA sequences. Nature **362:** 745–747.

Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in Drosophila populations. Annu. Rev. Genet. **11:** 49–78.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry: The Principles and Practice of Statistics in Biological Research*, Ed. 3. W. H. Freeman, New York.

Sperling, F. A. H., 1994 Sex-linked genes and species differences in Lepidoptera. Can. Entomol. **126:** 807–818.

Stuart, A., J. K. Ord and S. Arnold, 1999 *Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model*, Vol. 2A. Arnold, London.

Tan, C. C., 1946 Genetics of sexual isolation between *Drosophila pseudoobscura* and *Drosophila persimilis*. Genetics **31:** 558–573.

Tao, Y., D. L. Hartl and C. C. Laurie, 2001 Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proc. Natl. Acad. Sci. USA **98:** 13183–13188.

Tao, Y., Z-B. Zeng, J. Li, D. L. Hartl and C. C. Laurie, 2003 Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. Genetics **164:** 1399–1418.

Templeton, A. R., 1977 Analysis of head shape differences between two interfertile species of Hawaiian Drosophila. Evolution **31:** 630–642.

True, J. R., J. M. Mercer and C. C. Laurie, 1996a Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics **142:** 507–523.

True, J. R., B. S. Weir and C. C. Laurie, 1996b A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics **142:** 819–837.

Wu, C.-I, and A. W. Davis, 1993 Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. Am. Nat. **142:** 187–212.

Wu, C.-I, and H. Hollocher, 1998 Subtle is nature, pp. 339–351 in *Endless Forms: Species and Speciation*, edited by D. J. Howard and S. H. Berlocher. Oxford University Press, New York.

Val, F. C., 1977 Genetic analysis of the morphological differences between two interspecific species of Hawaiian Drosophila. Evolution **31:** 611–629.

Zeng, Z-B., J. Liu, L. F. Stam, C.-H. Kao, J. M. Mercer *et al.*, 2000 Genetic architecture of a morphological shape difference between two Drosophila species. Genetics **154:** 299–310.

Communicating editor: M. A. F. Noor

APPENDIX: STATISTICAL MODEL FOR THE ACCUMULATION OF HYBRID STERILITY

The model: We make the following assumptions:

1. The number of discrete HMS factors (N) on a unit length of chromosome (one polytene band here) has a Poisson distribution with mean λ:

$$f(N|\lambda) = e^{-\lambda}\lambda^N/N!.$$

2. The magnitude of effect of each factor (w) varies as an exponential distribution with a mean of $1/\theta$:

$$f(w|\theta) = \theta e^{-\theta w}.$$

3. The parameters λ and θ are homogeneous along the whole length of the chromosome (*i.e.*, no hot spots).

The cumulative effect (z) on a unit length of chromosome is

$$z = \sum_{k=1}^{N} w_k. \tag{A1}$$

The pdf and cdf of z can be derived as described below:

The moment-generating function of z is

$$\begin{aligned}\mathrm{mgf}_z(t) &= E[e^{tz}]\\ &= \sum_{N=0}^{\infty}\left\{\frac{\lambda^N}{N!}e^{-\lambda}\left[\int\cdots\int e^{t\left(\sum_{i=1}^{N} w_i\right)}\left(\theta^N e^{-\theta\sum_{i=1}^{N} w_i}\right)dw_1dw_2\ldots dw_N\right]\right\}\\ &= \sum_{N=0}^{\infty}\left[\frac{\lambda^N}{N!}e^{-\lambda}\left(\frac{\theta}{\theta-t}\right)^N\right] = \sum_{N=0}^{\infty}\left[\frac{(\lambda\theta/(\theta-t))^N}{N!}e^{-\lambda}\right]\\ &= e^{-\lambda}\left(1-\frac{1}{1-t/\theta}\right).\end{aligned} \tag{A2}$$

The pdf of z is derived by taking a result from Searle (1971, p. 49): For a noncentral χ^2 distribution $\chi^2(n, \lambda)$, the pdf is

$$f(u) = e^{-\lambda}\left[\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\frac{u^{(n/2)+k-1}e^{-u/2}}{2^{(n/2)+k}\,\Gamma((n/2)+k)}\right]$$

with a known mgf,

$$\mathrm{mgf}_u(s) = (1-2s)^{-n/2}\,e^{-\lambda(1-1/(1-2s))}. \tag{A3}$$

Comparing (A2) and (A3), let $n = 0$ and $s = t/2\theta$ and introduce a function $\delta(u)$ (Stuart *et al.* 1999, p. 243):

$$\delta(u) = \begin{cases}1, & u = 0\\ 0, & u \neq 0.\end{cases}$$

The pdf of u, where $u = 2\theta z$, is

$$\begin{aligned}f(u) &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\frac{u^{k-1}\,e^{-u/2}}{2^k\,\Gamma(k)}\\ &= e^{-\lambda}\frac{e^{-u/2}}{u}\left[\sum_{k=1}^{\infty}\frac{(\lambda u/2)^k}{k!(k-1)!}\right] + e^{-\lambda}\delta(u).\end{aligned}$$

Therefore, the pdf of z is

$$\begin{aligned}f(z|\theta,\lambda) &= 2\theta f(2\theta z)\\ &= e^{-(\lambda+\theta z)}\left[\sum_{k=1}^{\infty}\frac{(\lambda\theta)^k z^{k-1}}{\Gamma(k+1)\Gamma(k)}\right] + e^{-\lambda}\delta(z).\end{aligned} \tag{A4}$$

From (A2), the expectation and variance for z are λ/θ and $2\lambda/\theta^2$, respectively. As an illustration, some pdf's of (A4) are shown graphically (Figure A1). (A4) is essentially a summation of Poisson$(k|\lambda)$ × Gamma$(z|k,\theta)$ on k from 1 to infinity plus $e^{-\lambda}\delta(z)$:

$$f(z|\theta,\lambda) = \sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\frac{\theta^k z^{k-1}}{\Gamma(k)}\,e^{-\theta z} + e^{-\lambda}\delta(z).$$

In practice, we are more interested in the distribution of the hybrid sterility factors on the introgression with a length of L. For this, we substitute λ by λL and the distribution of z has expectation and variance of $\lambda L/\theta$ and $2\lambda L/\theta^2$, respectively.

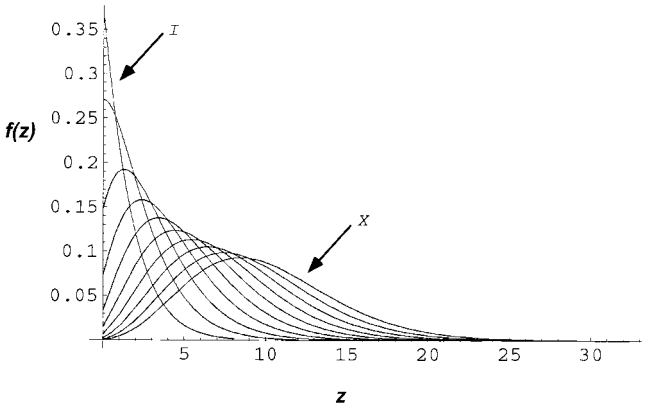

Figure A1.—An example for the pdf of z (Equation A4). The parameters are $\theta = 1.0$, $\lambda = 1.0$ (for I), . . . , $\lambda = 10.0$ (for X).

Proof: To prove that (A4) is the correct pdf of z, we use the Laplace transformation to show that $\mathrm{LT}[f(z)] =$ mgf of z, because by definition,

$$\mathrm{LT}[f(z)] = \int_0^\infty e^{-rz} f(z)\,dz = \int_0^\infty e^{tz} f(z)\,dz = E(e^{tz}).$$

We have the Laplace transformation

$$\mathrm{LT}\left[\frac{\theta^k u^{k-1} e^{-\theta u}}{\Gamma(k)}\right] = \left(\frac{\theta}{\theta - t}\right)^k.$$

Then,

$$\begin{aligned}\mathrm{LT}\left[e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\frac{\theta^k u^{k-1}e^{-\theta u}}{\Gamma(k)}\right] &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\mathrm{LT}\left[\frac{\theta^k u^{k-1}e^{-\theta u}}{\Gamma(k)}\right]\\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\left(\frac{\theta}{\theta - t}\right)^k\\ &= e^{-\lambda+\lambda\theta/(\theta-t)}\left[\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\left(\frac{\theta}{\theta - t}\right)^k e^{-\lambda\theta/(\theta-t)}\right]\\ &= e^{-\lambda+\lambda\theta/(\theta-t)}\,(1 - e^{-\lambda\theta/(\theta-t)}) = e^{\lambda t/(\theta-t)} - e^{-\lambda}\end{aligned}$$

and

$$\mathrm{LT}[e^{-\lambda}\delta(u)] = e^{-\lambda}.$$

Therefore, from (A4) we have

$$\mathrm{LT}[f(z)] = e^{\lambda t/(\theta-t)},$$

which is (A2), therefore proving that (A4) is the pdf of z.

If we assume that w has a Gamma distribution,

$$f(w|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, w^{\alpha-1}e^{-\beta w}$$

by the same derivation from (A1) to (A4), we have the distribution for z:

$$f(z|\theta, \lambda) = e^{-(\lambda+\beta z)}\left[\sum_{k=1}^{\infty}\frac{(\lambda\beta^\alpha)^k z^{\alpha k-1}}{\Gamma(k+1)\Gamma(\alpha k)}\right] + e^{-\lambda}\delta(z). \tag{A5}$$

We note that (A5) is essentially the same model as that intuitively proposed to account for mutation accumulation along a chromosome (Keightley and Ohnishi 1998, Equation 2). For this distribution, $E(z) = \alpha\lambda/\beta$ and $\mathrm{Var}(z) = \lambda\alpha(1+\alpha)/\beta^2$. If $E(z)$ and $\mathrm{Var}(z)$ can be estimated from mutation accumulation, the number of mutations (U) and their mean effect ($\bar{w}$) can be obtained,

$$U = \lambda = \frac{1+\alpha}{\alpha}\frac{E(z)^2}{\mathrm{Var}(z)} \quad \text{and} \quad \bar{w} = \frac{\alpha}{\beta} = \frac{\alpha}{1+\alpha}\frac{\mathrm{Var}(z)}{E(z)}.$$

It is clear that $U\bar{w} = E(z)$. However, without knowledge of the shape (defined by α) of the Gamma distribution of z, U and $\bar{w}$ cannot be estimated reliably. For the assumption of equal effects for z ($\alpha \to \infty$) (*e.g.*, Bateman 1959; Mukai *et al.* 1972), $U \geq E(z)^2/\mathrm{Var}(z)$ and $\bar{w} \leq \mathrm{Var}(z)/E(z)$, which are the upper and lower bounds for the number of mutations and their average effect, respectively. For a more realistic assumption that the mutation effect has an exponential distribution ($\alpha = 1$), the two estimates are

$$U = 2\frac{E(z)^2}{\mathrm{Var}(z)} \quad \text{and} \quad \bar{w} = \frac{1}{2}\frac{\mathrm{Var}(z)}{E(z)}.$$

Because α is unknown, the maximum-likelihood estimates with an assumption of $w \sim \mathrm{Gamma}(w|\alpha, \beta)$ (Keightley 1996) do not appear to be more reliable than those of the Bateman-Mukai method. Recent numerical studies have confirmed this conclusion (Keightley 1998; Fry *et al.* 1999).