Rapid Evolution of a Coadapted Gene Complex: Evidence From the *Segregation Distorter* (*SD*) System of Meiotic Drive in *Drosophila melanogaster*

Michael F. Palopoli and Chung-I Wu¹

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

Manuscript received October 2, 1995

Accepted for publication April 25, 1996

ABSTRACT

Segregation Distorter (*SD*) is a system of meiotic drive found in natural populations of *Drosophila melanogaster*. Males heterozygous for an *SD* second chromosome and a normal homologue (*SD*⁺) produce predominantly *SD*-bearing sperm. The coadapted gene complex responsible for this transmission advantage spans the second chromosome centromere, consisting of three major and several minor interacting loci. To investigate the evolutionary history of this system, we surveyed levels of polymorphism and divergence at six genes that together encompass this pericentromeric region and span seven map units. Interestingly, there was no discernible divergence between *SD* and *SD*⁺ chromosomes for any of these molecular markers. Furthermore, *SD* chromosomes harbored much less polymorphism than did *SD*⁺ chromosomes. The results suggest that the *SD* system evolved recently, swept to appreciable frequencies worldwide, and carried with it the entire second chromosome centromeric region (roughly 10% of the genome). Despite its well-documented genetic complexity, this coadapted system appears to have evolved on a time scale that is much shorter than can be gauged using nucleotide substitution data. Finally, the large genomic region hitchhiking with *SD* indicates that a multilocus, epistatically selected system could affect the levels of DNA polymorphism observed in regions of reduced recombination.

Clearly a higher plant species is at the mercy of its pollen grains. A gene which greatly accelerates pollen tube growth will spread through a species even if it causes moderately disadvantageous changes in the adult plant. J. B. S. HALDANE (1932)

THE concept of the “selfish gene” was introduced by the architects of the modern synthesis (FISHER 1930; WRIGHT 1931; HALDANE 1932). Indeed, the notion that genes are competing with rival alleles for representation on the chromosomes of future generations is fundamental to the reconciliation of Mendelian heredity with the theory of evolution by natural selection. As was recognized long ago by HALDANE (1932, quoted above), selection will tend to favor a gene that enhances the success of the *gametes* it inhabits, and this can remain true even if this same gene also poses a significant fitness cost during the diploid portion of the life cycle. The *Segregation Distorter* (*SD*) system of meiotic drive—arguably the best-understood example of gametic transmission ratio distortion—behaves much as HALDANE envisioned: Although *SD* enhances the success of those male gametes it inhabits (HARTL *et al.* 1967), it also poses large fitness costs to diploid individuals of both sexes (NICOLETTI *et al.* 1967; WU *et al.* 1989). This selfish genetic element is found at frequencies of 1–5% in most natural popula-

tions of *Drosophila melanogaster* (HIRAIZUMI and NAKAZIMA 1965; HARTL and HIRAIZUMI 1976; TEMIN and MARTHAS 1984).

Males heterozygous for an *SD* second chromosome and a normal homologue (*SD*⁺) usually produce more than 90% *SD*-bearing sperm; the *SD* chromosome gains this transmission advantage by inducing the dysfunction of rival spermatids that received the *SD*⁺ chromosome (HARTL *et al.* 1967; NICOLETTI *et al.* 1967; TOKUYASU *et al.* 1977; HAUSCHTECK-JUNGEN and HARTL 1982). The *SD* complex straddles the second chromosome centromeric region and involves three major, tightly linked loci: *Segregation distorter* (*Sd*), located in the euchromatin near the base of 2L, encodes a *trans*-acting factor that induces dysfunction of those spermatids containing sensitive alleles at the *Responder* locus; *Enhancer of Segregation Distorter* [*E(SD)*], located in the centric heterochromatin of 2L, is required for the full expression of drive and is also a weak driver in its own right; and *Responder* (*Rsp*), an array of satellite DNA embedded deep within the heterochromatin at the base of 2R, not only serves as a target site for the action of *Sd* but appears to have a positive function in the soma as well (GANETZKY 1977; BRITTNACHER and GANETZKY 1983, 1984; WU *et al.* 1988; WU *et al.* 1989; TEMIN 1991; TEMIN *et al.* 1991). Recent molecular and genetic investigations into the major components of the *SD* complex led to the successful cloning of both *Sd* and *Rsp* and paved the way for the work reported here (WU *et al.* 1988; POWERS and GANETZKY 1991; MCLEAN *et al.* 1994; PALOPOLI *et al.* 1994).

Corresponding author: Michael F. Palopoli, Howard Hughes Medical Institute, MC 1028, AMB N-101, 5841 South Maryland Ave., University of Chicago, Chicago, IL 60637.
E-mail: palo@midway.uchicago.edu

¹ E-mail: ciwu@pondside.uchicago.edu

An investigation into the molecular population genetics of this system is important for at least two reasons. First, because of the extensive genetical analyses that have been done, we understand some of the strong deterministic forces that are expected to influence its evolution. For example, the epistatic nature of the entire *SD* gene complex leads us to expect strong patterns of linkage disequilibrium, particularly in the region from *Sd* to *Rsp* (WU and HAMMER 1991). This is because recombination between *SD* and *SD*⁺ chromosomes in this region will result in two chromosomes doomed to rapid extinction: one "suicide" chromosome that self-distorts in males (HARTL 1974), and one chromosome bearing a deleterious *Rsp* array but lacking meiotic drive (CHARLESWORTH and HARTL 1978; WU *et al.* 1989). Indeed, regardless of the presence or absence of inversions, all *SD* chromosomes isolated from nature have been found to carry *Sd*, *E(SD)*, and extremely insensitive *Rsp* alleles, whereas non-*SD* chromosomes typically carry wild-type alleles at all three loci (SANDLER and GOLIC 1985; TEMIN *et al.* 1991). The question is to what extent this statistical association at the functional level, apparently caused by strong epistatic selection, is reflected in patterns of association among linked DNA polymorphisms.

Second, DNA-polymorphism data for both *SD* and *SD*⁺ chromosomes in a region known to be tightly linked to the *Sd* locus will shed light on the evolutionary history of the meiotic drive system itself. This is particularly interesting given the well-documented complexity of the system: In addition to the three major loci described above, there are several other linked drive enhancers, numerous unlinked drive suppressors, and a variety of *SD*-chromosome inversion patterns (HARTL and HIRAZUMI 1976; LYTTLE 1991, 1993) found in natural populations. We would like to determine the approximate time scale involved in the evolution of this coadapted gene complex. Results for *SD* will also serve as a comparison to what we know about the evolution of two other meiotic drive systems: the *t* haplotypes in house mice (*e.g.*, WU and HAMMER 1991) and the Sex-Ratio chromosomes in *D. pseudoobscura* (BABCOCK and ANDERSON 1996).

To address these issues, we surveyed *SD* and *SD*⁺ chromosomes for levels of polymorphism and divergence at a series of six loci encompassing the second chromosome centromeric region (see Figure 1, proceeding left to right): Alcohol dehydrogenase (*Adh*; 2-50.1), Dopa decarboxylase (*Ddc*; 2-53.9), Topoisomerase 2 (*Top2*; 2-54), *light* (*lt*; 2-55), *Responder* (*Rsp*; 2-55), and Glyceraldehyde phosphatase dehydrogenase-1 (*Gapdh-1*; 2-57). These loci were chosen for their strategic locations with respect to the *SD* system: the *Top2* gene is immediately proximal and physically adjacent to the *Sd* locus (POWERS and GANETZKY 1991; MCLEAN *et al.* 1994; PALOPOLI *et al.* 1994), so the evolutionary history at *Top2* is expected to reflect accurately that of *Sd*; similarly, the *lt*

locus is located in the same stretch of β -heterochromatin as the *E(SD)* locus (BRITTNACHER and GANETZKY 1984), and given the extremely low rates of recombination in heterochromatin (ASHBURNER 1989 and references therein), the evolutionary history determined for the *lt* locus region should reflect accurately that of *E(SD)*; the *Rsp* locus, as the target of distortion, is itself one of the major components of the gene complex and so should also reflect the history of the *SD* system; *Adh* and *Ddc* are located at different distances from the *Sd* locus proceeding distally onto 2L (refer to Figure 1), thus providing information on the degree of linkage disequilibrium between the *SD* complex and this euchromatic region; and, finally, *Gapdh-1* is located in the euchromatin near the base of 2R, thus providing comparable information on the degree of linkage disequilibrium between the *SD* complex and this second euchromatic region. Allozymes (*Adh*), DNA sequences (*Top2* and *Rsp*), restriction fragment length polymorphisms (*Ddc*, *lt*, and *Rsp*), and single-strand conformation polymorphisms (*Gapdh-1*) were all used in this survey.

Our results suggest that a complex meiotic drive system such as *SD* can evolve so rapidly that there is no detectable neutral divergence between *SD* and *SD*⁺ haplotypes. Furthermore, the results indicate that a meiotic drive system—and supposedly other multilocus, epistatically selected gene complexes—could have a profound impact on levels of polymorphism and patterns of linkage disequilibrium in regions of reduced recombination (see BEGUN and AQUADRO 1992; AQUADRO and BEGUN 1993).

MATERIALS AND METHODS

Genomic DNA was prepared and Southern blots were carried out using standard protocols adapted from SAMBROOK *et al.* (1989).

Wild-type chromosomes: The following *SD*⁺ *Drosophila melanogaster* isofemale lines were provided by R. TEMIN and chosen (arbitrarily) for *Top2* sequence analysis: Australia 86-5, Cam 108 (North Carolina), Canton S (Ohio), France V2-1, Japan QD18, LemonCove (California), M113 (Wisconsin), and Oregon R. The *D. simulans* isofemale lines analyzed at *Top2* were provided by the Indiana University Stock Center, as follows: Sim911 (Guatemala), Sim912 (Kushla, Alabama), Sim913 (Morro Bay, California), Sim914 (South Africa), and Sim-yvf (unknown). R. TEMIN also provided *SD*⁺ second chromosome lines of *D. melanogaster* that were extracted from populations in Sonoma County, California (designated "C" followed by a number), and Madison, Wisconsin (designated "M" or "MM" followed by a number), as follows: C25-3, C38, C46, C55, C83, C100, C111, C115, C171, C182, M10, M32, M65, M74, M87, M93, M107, M113, M881, M1115, MM27, MM33, MM153, and MM204 (refer to TEMIN and MARTHAS 1984 for more information on these chromosomes). Ten additional *SD*⁺ second chromosome lines were collected from a single bait at the Closs Farms orchard, near Rochester, New York, in 1988 (designated "CL" followed by a number). All wild-type chromosomes (though not all for every locus) were used for restriction fragment analysis at the *Rsp*, *lt*, and *Ddc*

loci, as well as single-strand conformational polymorphism (SSCP) analysis at *Gapdh-1*.

SD chromosomes: We analyzed polymorphisms for the following SD chromosomes (from natural populations in Madison, Wisconsin unless otherwise noted): *SD-5*, *SD-25-1*, *SD-72*, *SD-79bw*, *SD-146-1B*, *SD-249-1C*, *SD-258A*, *SD-1462A*, *SD-Arm28* (Armidale, Australia), *SD-Arm1* (Armidale, Australia), *SD-C11*, *SD-LA* (Los Arenos, Spain), *SD-Mad*, *SD-Roma* (Rome, Italy), *SDT87*, *SD-V017B* (Spain), and *SD-Wey* (Rhode Island). These stocks were provided by T. LYTLE and R. TEMIN and are described in greater detail elsewhere (HARTL and HIRAZUMI 1976; TEMIN and MARTHAS 1984). Not all SD chromosomes were surveyed at all loci.

Genetic stocks and crosses: For a complete description of the stocks used see LINDSLEY and ZIMM (1992). All second chromosomes analyzed in this study that are not viable and fertile when homozygous were maintained in heterozygous state with the multiply-inverted balancer chromosome *CyO* in order to prevent recombination. To examine polymorphisms in the region of the *Sd* locus, individuals heterozygous for the chromosome of interest were mated to *Df(2L)TW158/CyO* (this deficiency includes polytene bands 37B2-8; 37E2-F4, encompassing the *Ddc*, *Sd*, and *Top2* loci) and the resulting straight-winged progeny were maintained as a stock. To study *Rsp* restriction fragments, SD chromosomes were made heterozygous with a deficiency for most of the *Rsp* array [*Df(2R)Rsp¹⁶cn bw*, designated *R16*]. Recombination between *Rsp* and *cn* allowed the ratio of *Rsp* alleles in each stock to vary from 1:1, making relative band intensities uninformative in subsequent Southern blots. To verify that the SD stocks chosen had not been contaminated by an *SD⁺* homologue, we measured the ability of each to distort a sensitive chromosome (data not shown) and also probed DNA from that stock with a 6.7-kb *EcoRI* fragment of the *Sd* locus (provided by P. POWERS; see POWERS and GANETZKY 1991; MCLEAN *et al.* 1994) to verify that the chromosome in question carried the SD-specific tandem duplication (data not shown).

Sequencing strategy: A 1.1-kb fragment of *Top2* was PCR-amplified using 24-mer oligonucleotide primers starting at positions 372 (+ primer 5' base) and 1486 (– primer 5' base) in the sequence reported by WYCKOFF *et al.* (1989). *Top2* is located immediately proximal to the SD-specific tandem duplication (POWERS and GANETZKY 1991), which has been determined to be the *Sd* locus itself (MCLEAN *et al.* 1994; PALOPOLI *et al.* 1994). The *Top2* polyadenylation site is approximately 180 bp upstream of the proximal end of the 5-kb tandem duplication. For *Rsp* sequencing, oligonucleotide primers starting at positions 84 (+ primer 5' base, 19 bp) and 86 (– primer 5' base, 23 bp) in the *Rsp* consensus sequence reported by WU and HAMMER (1991) were used to amplify *Rsp* repeats from *SD-Mad* homozygotes (see below). All amplifications were carried out in 50 μ l volumes with 1 unit of *Taq* polymerase (Amersham) using the reaction buffer supplied with the enzyme. The *Top2* PCR products were digested with *BglII* and *PstI* and then electrophoresed on an agarose gel; the resulting 0.5-kb band was cut out of the gel and purified with the GeneClean kit (Bio 101) and then cloned into M13-mp18 and -mp19 for bidirectional sequencing. The *Rsp* PCR products were blunt-end ligated into M13-mp18. Cloned products were dideoxy-sequenced with Sequenase (Stratagene). In the case of *Top2*, between three and five clones were sequenced in each direction in order to avoid mistaking *Taq*-polymerase incorporation errors for actual base substitutions. Because the *Rsp* PCR product was amplified from a heterogeneous assemblage of different *Rsp* repeats, each of the five clones was assumed to represent the sequence of a different original repeat (this will result in *Taq* polymerase

errors going undetected, so the divergence between *Rsp* repeats from *SD-Mad* vs. *SD⁺* will tend to be overestimated).

Sequences were aligned both manually and with the assistance of the "Gap" program of the UWGCG package. Estimates of sequence differences were calculated using a two-parameter correction for noncoding sequences (KIMURA 1980). Divergence between *D. melanogaster* and *D. simulans* sequences was compared to results for other euchromatic loci for these species in order to rule out evolutionary scenarios invoking strong selection or unusual mutation pressures acting on the intron sequences. For purposes of estimating θ using *Top2* in *D. melanogaster*, the five SD sequences were considered a sample of one gene. GenBank accession numbers for DNA sequences described in this paper are as follows: U53787–U53804 (*Top2*) and U53805–U53809 (*Rsp*).

***Rsp* restriction fragment analysis:** The *Rsp* arrays vary tremendously in size and organization (WU *et al.* 1988; CABOT *et al.* 1993), and they are interspersed with an unknown number of other repetitive sequences (TEMIN *et al.* 1991). We do not know what effects variation in these various factors might have upon the intrinsic mutation rates of a given array. Furthermore, it is likely that powerful selective forces are acting upon *Rsp* arrays of different sizes (see CHARLESWORTH and HARTL 1978; WU *et al.* 1989). Because of these complexities, only a qualitative analysis of the *Rsp* RFLP data is presented here.

RFLPs at *lt*: We used the complete *lt* locus cDNA (provided by B. T. WAKIMOTO) to probe Southern blots of genomic DNA that had been digested with one of four six-cutter restriction enzymes (*Bam*HI, *Bgl*II, *Eco*RI, and *Xba*I). The *lt* exons are interspersed with more than 10 kb of middle repetitive DNA, most of which is thought to consist of transposable element insertions (DEVLIN *et al.* 1990). This makes mapping individual restriction sites difficult. Instead, the resulting restriction fragment lengths were just used to distinguish allelic classes and determine their frequencies among both SD and *SD⁺* chromosomes. A list of the restriction fragment lengths associated with the various allelic classes is available upon request from M. F. PALOPOLI. Any homozygous-lethal chromosomes were first made heterozygous with a deficiency for this region (*Df_{ltx5}*, provided by B. WAKIMOTO; DEVLIN *et al.* 1990; WAKIMOTO and HEARN 1990).

Allozyme polymorphism at *Adh*: To study allozymes at *Adh*, individuals heterozygous for SD were crossed to *WA-Slow* homozygotes and the resulting SD heterozygous individuals were used for protein electrophoresis. Cellulose acetate gel electrophoresis, performed according to HEBERT and BEATON (1989), was used to distinguish allozymes.

RFLPs at *Ddc*: Taking advantage of previous survey results for 46 *SD⁺* chromosomes (AQUADRO *et al.* 1992), we sampled 16 SD chromosomes and eight (additional) *SD⁺* chromosomes at two sites known to be highly polymorphic in natural populations (*i.e.*, *Bgl*II, –29.8 and *Bam*HI, +12.0 in AQUADRO *et al.* 1992). The presence/absence of these restriction sites resulted in four possible two-site haplotypes, and the combined data were used for comparing SD and *SD⁺* haplotype frequency distributions.

SSCPs at *Gapdh-1*: Based on an unpublished sequencing study of 11 *SD⁺* alleles conducted by J. McDONALD and M. KREITMAN, it was determined that there are two common haplotypes at *Gapdh-1* that can be distinguished from each other by two tightly linked substitutions in the first intron of this gene. A pilot study demonstrated that this polymorphism is readily detected using the SSCP technique (data not shown; ORITA *et al.* 1988), allowing for an efficient survey of a collection of SD vs. *SD⁺* chromosomes for the relative frequencies of these two common haplotypes.

FIGURE 1.—Diagram of the approximate map positions of the genes surveyed (enclosed in black rectangles), relative to some of the components of the *SD* gene complex, as well as the second chromosome centromeric region. At the top is a diagram of the entire second chromosome, with the centromere depicted as a black oval, the heterochromatic regions as shaded rectangles, and the euchromatic arms as black lines. Below is a magnified view of the map positions for the region surrounding the centromere. Note that the entire heterochromatic region, including the centromere, is “squeezed” into a region of unknown recombination size at map position 55, as there is almost no documented recombination within the heterochromatin (ASHBURNER 1989 and references therein). Both *lt* and *E(SD)* are situated in the β -heterochromatin of 2L, whereas *Rsp* is the most proximal known locus in the α -heterochromatin at the base of 2R. *M(SD)* is a linked enhancer of segregation distortion.

RESULTS

The results are organized according to the type of insight each gene provided into the evolutionary history of the *SD* complex (refer to Figure 1). First, we present the data that were most informative for each of the three major component loci separately. The major loci of the *SD* complex are *Sd*, *Rsp*, and *E(SD)*, which correspond to the results presented below for *Top2*, *Rsp*, and *lt*, respectively. Next, analyses of linked euchromatic regions are presented, as follows: *Adh* and *Ddc* provided measures of linkage disequilibrium at two distances from the *SD* complex proceeding distally on 2L; similarly, *Gapdh-1* provided a measure of disequilibrium at the euchromatic base of 2R, as well as a measure of *SD*-*SD*⁺ divergence in the vicinity of a locus that enhances segregation distortion (*Modifier of SD*; HIRAZUMI *et al.* 1980).

***Sd* locus region:** We sequenced a fragment of *Top2*, which is immediately proximal to the *Sd* locus, for eight *SD*⁺ and five *SD* chromosomes from *D. melanogaster*, five second chromosomes from *D. simulans*, and one from *D. sechellia*. In *D. melanogaster*, although there was slight variation in length due to small insertions/deletions, 538 bp were sequenced on the average—472 bp from the three-prime half of the large first intron and 66 bp from the second exon. In *D. simulans*, the first intron portion of the fragment varied somewhat more in length, from about 420 to 460 bp, depending on the allele sequenced. In the interests of examining a collection of sites under relatively uniform evolutionary forces and unlikely to be under selection directly, only the intron portion of the sequence was used for analysis. An alignment of multiclonal consensus sequences is presented in APPENDIX A.

The levels of DNA polymorphism and divergence among wild-type chromosomes were comparable to results for other euchromatic loci in these species (AQUADRO and BEGUN 1992; KREITMAN and WAYNE 1994). Within *D. melanogaster*, considering the five sequences

obtained from *SD* chromosomes to represent a sample of one because they could not be considered independent isolates from nature, the average number of nucleotide differences per site between two sequences (nucleotide diversity) was 0.0035 ± 0.00086 (mean and standard error calculated following NEI 1987). Within *D. simulans*, the average nucleotide diversity was 0.0412 ± 0.01158 . Estimates of θ ($= 4N\mu$) calculated using the number of segregating sites (WATTERSON 1975) were 0.0039 for *D. melanogaster* and 0.0424 for *D. simulans*. The average divergence between *D. melanogaster* and *D. simulans* sequences, corrected for multiple hits by the two-parameter method (KIMURA 1980) but not corrected for polymorphism within species, was 0.1298.

The five sequences sampled from *SD* chromosomes were identical to each other and closely related to wild-type *D. melanogaster* chromosomes. In fact, they were identical to one wild-type sequence (Aus 86-5). The combined *SD* and *SD*⁺ results can be summarized in the form of an unrooted tree, produced by the neighbor-joining algorithm (SAITOU and NEI 1987), depicting the evolutionary relationships among sequences (Figure 2). Whereas the branch lengths represent relatively “normal” levels of polymorphism and divergence among wild-type sequences for these species (*e.g.*, KREITMAN and WAYNE 1994), the *Top2* sequences from *SD* chromosomes are shown to be monomorphic and not diverged from *SD*⁺ sequences.

Regarding the *SD* system, two questions can be asked of these data. First, is there significant linkage disequilibrium between *SD* allelic state and the haplotype at *Top2*. A standard test for nonrandom gametic association between alleles of two genes is to construct a 2×2 table of allele counts and conduct a test of independence (*e.g.*, HARTL and CLARK 1989). Here, the observed *Top2* sequence haplotypes were divided into two allelic classes: those exhibiting the same sequence as the first *SD* chromosome sequenced *vs.* all other sequences. Because the 5 *SD* chromosomes were identical in se-

FIGURE 2.—Neighbor-joining tree depicting the relationships among the *Top2* DNA sequences. The specific topology of this tree is not important for the arguments presented in this report. Note, instead, the levels of polymorphism and divergence among *SD*⁺ vs. *SD* haplotypes. Whereas the wild-type alleles exhibit the expected levels of polymorphism and divergence for noncoding DNA in a euchromatic region in these species, all five *SD* sequences are identical to one *SD*⁺ sequence in *D. melanogaster* (Aus 86-5). The probability of this occurring by chance alone, assuming *SD* and *Top2* were actually in linkage equilibrium, is approximately 0.0048 ± 0.0001 .

quence, the counts for the first row of such a table are 5 and 0; because only one of the eight *SD*⁺ alleles was identical to the *SD* sequences, the counts for the second row are 1 and 7. Using Fisher's exact test for independence, the probability of observing this distribution of *Top2* haplotype counts due to chance alone (if these genes were actually in linkage equilibrium) is $P = 0.0047$. We concluded that there is significant linkage disequilibrium between *Sd* and *Top2*.

Second, how old is the *SD* system according to these data? The simplest answer is that it cannot be very old at all since one *SD*⁺ allele (Aus 86-5) shared exactly the same sequence as the five *SD* chromosomes. To address this question more quantitatively, we tabulated on the one hand the divergence between the identical *SD* sequences and each *SD*⁺ sequence, and on the other hand the divergence among *SD*⁺ sequences. The averages were then compared, as follows: the average divergence between *SD*-*SD*⁺ was 0.0015 and between *SD*⁺-*SD*⁺ was 0.0034. If *SD*-*SD*⁺ was greater than *SD*⁺-*SD*⁺ divergence, then we could assess the significance of the difference (using the Mann-Whitney *U* test, for example); however, given that the average of observed divergence between *SD*-*SD*⁺ was actually less than between *SD*⁺-*SD*⁺, the results of such a test would be difficult to interpret (the confusing fact that the average *SD*-*SD*⁺ divergence was less than that among *SD*⁺ alleles can be understood by examining Figure 2 and noting that the *SD* sequence

FIGURE 3.—Southern blot of the *Rsp* arrays for the designated *SD*⁺ isochromosomal stocks. Genomic DNA was prepared from isochromosomal stocks and digested with *Msp*I. Note that every restriction fragment pattern shown is different from all others, illustrating the extremely high levels of diversity that is typically found among *Rsp*⁺ haplotypes.

is located "in the middle" of the two extremes among the *SD*⁺ sequences).

Finally, in order to corroborate the lack of sequence divergence between *SD* and *SD*⁺, we sequenced an additional 421 bp (162 bp from the 3' untranslated region and 265 bp from the 5' half of the *Top2* first intron) from one *SD* chromosome (*SD*-*Mad*) and compared it to the published *SD*⁺ (Canton-S) sequence (data not shown; WYCKOFF *et al.* 1989; PALOPOLI 1995). There were a total of two differences between *SD*-*Mad* and Canton-S for these 421 bp, which gives an estimate of nucleotide diversity (0.0048) that agrees closely with the results presented above.

***Rsp* locus variation:** We surveyed RFLPs and DNA sequences at the *Rsp* locus. Like many families of tandem-repetitive DNA, *Rsp* exhibits a great deal of RFLP variation among chromosomes (WU *et al.* 1988; WU and HAMMER 1991). For example, the RFLP patterns of 10 *SD*⁺ chromosomes collected together at the same bait were all distinct [Figure 3; see DOSHI *et al.* (1991) and CABOT *et al.* (1993) for additional details concerning large-scale structure at *Rsp*]. In contrast to the high degree of variability exhibited by *SD*⁺ chromosomes, all *SD* chromosomes that were examined exhibited one of two similar RFLP patterns (Figure 4). Both RFLP patterns suggested just small blocks of *Rsp* repeats, which is consistent with the extreme insensitivity of the *Rsp* alleles on *SD* chromosomes and the strong correlation between *Rsp* array size and sensitivity to distortion (WU *et al.* 1988). One of the two blocks appeared to be shared among all *SD*s because that band was shared by the two distinguishable *SD* haplotypes across several restriction enzymes (Figure 4, and additional data not

FIGURE 4.—Southern blots of the *Rsp* arrays for the designated *SD* genotypes. Note that *R16* and the first three *SD* chromosomes are homozygous; the remainder are heterozygous mixtures of *SD* and *R16* (because the ratio of *SD*:*R16* was not maintained at 1:1, the relative intensities of bands cannot be considered informative). Genomic DNA was digested with *AluI* (top), *MspI* (bottom), and several other enzymes (data not shown). Note that for all enzymes examined there were only two *Rsp* haplotypes among *SD* chromosomes (the “*SD-Mad* type” shared by a total of three *SD* chromosomes, and the “*SD-V017* type” shared by 9 *SD* chromosomes). Furthermore, the restriction fragment patterns suggest that the two haplotypes harbor qualitatively similar *Rsp* arrays (*i.e.*, small blocks of *Rsp* repeats, one of which is shared between haplotypes).

shown). Note that we do not attach significance to the fact that bands are shared between the *SD* chromosomes and *R16* (Figure 4), since it has been reported that a small number of *Rsp* repeats actually reside in the centric heterochromatin of the third chromosome (R. CAIZZI, C. CAGGESE and S. PIMPINELLI, unpublished results).

In addition, we PCR-amplified, cloned, and sequenced five stretches of *Rsp* DNA from one *SD*-chromosome (*SD-Mad*), and these sequences are shown after alignment with published *Rsp* repeat sequences in APPENDIX B. The *Rsp* locus harbors two repeat families that have been designated “left” and “right” repeats based on the dimeric organization of the *Rsp* array (WU *et al.* 1988; WU and HAMMER 1991). There was no appreciable divergence between repeats sampled from *SD-Mad* and those that have been published for *SD*⁺ chromosomes (this can be seen by inspection of APPENDIX B). The average divergence between “left” repeats from *SD-Mad* *vs.* *SD*⁺ chromosomes, for a combined total of 326 bp, was 0.071; the average divergence between “right” repeats from *SD-Mad* *vs.* *SD*⁺ chromosomes, for a combined total of 226 bp, was 0.025. So the average divergence between repeats from *SD-Mad* *vs.* *SD*⁺ re-

peats from within the same repeat family was 0.052. This value falls well within the range observed for comparisons within and between arrays on wild-type chromosomes (see Figure 3 in CABOT *et al.* for the distribution of these values). It is possible that the repeats we sequenced were actually located in the third chromosome centromeric region (R. CAIZZI, C. CAGGESE and S. PIMPINELLI, unpublished results); this seems unlikely, however, given the high level of similarity to the sequences published for other second-chromosome *Rsp* arrays (*i.e.*, a nonhomologous array of repeats should exhibit greater average divergence, unless the *Rsp* duplication was very recent). Furthermore, the interpretation that all five sequences were from the third chromosome is not a parsimonious one. It requires not only that *SD* chromosomes do not contain *Rsp* sequences, but also that the translocation between 2 and 3 occurred recently. Unless further information on the *Rsp* sequences of chromosome 3 suggests otherwise, our interpretation is a more parsimonious one. We concluded that the *SD* chromosomes do not appear to exhibit any appreciable neutral divergence from *SD*⁺ chromosomes at *Rsp*, which is itself one of the major elements of the *SD* system.

***E(SD)* locus region:** Sixteen *SD* and 35 *SD*⁺ alleles were surveyed for RFLPs around the *lt* gene, which is located in the same stretch of β -heterochromatin as *E(SD)* at the base of 2L (see Figure 1; DEVLIN *et al.* 1990; WAKIMOTO and HEARN 1990). Altogether, 17 haplotypes were distinguished using four different six-cutter restriction enzymes. These alleles probably differed, for the most part, by transposable element insertions/deletions (DEVLIN *et al.* 1990; WAKIMOTO and HEARN 1990), but the relative frequency of insertions/deletions *vs.* restriction site changes was not determined. Among the 35 *SD*⁺ genes surveyed, one allele was observed 11 times, one allele was observed seven times, one was observed three times, one two times, and 12 alleles were observed only once. Interestingly, 15 out of the 16 *SD* chromosomes surveyed exhibited the same haplotype as the *most common allele* among *SD*⁺ genes; only *SD-V017B*, which is a naturally occurring mutant at the *lt* locus, exhibited a different (and unique) haplotype. Hence, we concluded that there is no detectable divergence between *SD* and *SD*⁺ chromosomes at *lt*.

The *SD*⁺ haplotype frequencies were compared to the neutral expectation using the statistical test of FU and LI (1993) with no outgroup. For this test, we considered each distinct *lt* haplotype to represent one mutational difference that has arisen during the phylogeny shared by these alleles; using this approach for the 35 *SD*⁺ genes that were surveyed, the resulting test statistic was $D^*[SD^+] = -3.42$ ($P < 0.01$). We concluded that the distribution of *lt* haplotype frequencies among *SD*⁺ chromosomes is inconsistent with the neutral prediction; the observed deviation was in the direction of “too many” singletons, which is consistent with a number

of evolutionary scenarios, including a recent selective sweep due to the *SD* complex or purifying selection against transposable element insertions [see FU and LI (1993) for a discussion of alternatives].

Regarding the extent of linkage disequilibrium between *SD* and *lt* allelic states, the data were analyzed in a manner identical to that described for the *Top2* significance test (above). First the haplotypes were collapsed into two categories: identical to *SD-Mad* vs. different than *SD-Mad*; this resulted in a 2×2 table with the values 16 and 1 for *SD* chromosomes, 11 and 24 for *SD*⁺ chromosomes. Based on Fisher's exact test for independence, the probability of observing this distribution of haplotype counts, assuming that *SD* and *lt* are actually in linkage equilibrium, is $P = 0.000015$. (If the data were not collapsed into two allelic classes, but a test of independence was instead conducted across all 17 allelic classes by a standard Monte Carlo permutation procedure, the estimated significance level is $P = 0.0099$; see results for *Ddc*, below, for additional details concerning this test procedure.) We concluded that the *SD* haplotypes are nonrandomly associated with the alleles at the *lt* locus.

Near the middle of 2L: We used cellulose acetate electrophoresis (HEBERT and BEATON 1989) to examine *Adh* allozymes for 16 *SD* chromosomes: 10 exhibited the *Adh*^S and six the *Adh*^F allele. Since the collection locations for these *SD* chromosomes were known, it was possible to calculate the expected frequencies (under the assumption of random association between *SD* and *Adh* allelic states) using the extensive data available for *Adh* allozyme frequencies worldwide (e.g., OAKESHOTT *et al.* 1982). Based on these data, we expected to observe 7.4 *Adh*^S and 8.6 *Adh*^F alleles among 16 chromosomes drawn at random from the corresponding locations. Even if we rounded the expected values so as to artificially enhance the observed level of significance (i.e., assume that, based on allozyme frequencies at the sampling locations, seven slow and nine fast alleles were expected), the observed allozyme frequencies were not significantly different from expected (Fisher's exact test, $P > 0.2$ in this case). Hence, we accepted the null hypothesis that *SD* and *Adh* are randomly associated (i.e., in linkage equilibrium).

Distal of *Sd* near the base of 2L: Sixteen *SD* alleles and eight *SD*⁺ alleles (in addition to the much larger sample of *SD*⁺ alleles that was already available in the literature) were surveyed for restriction fragment length polymorphisms (RFLPs) around the *Ddc* gene, which is approximately 0.1 cM distal from the *Sd* locus on 2L (refer to Figure 1; LINDSLEY and ZIMM 1992). Because a large survey of RFLPs among *SD*⁺ chromosomes had already been done (AQUADRO *et al.* 1992), this information could be used to direct the survey of *SD* chromosomes. In particular, by looking at sites that are known to be polymorphic and at intermediate frequencies among wild-type alleles, the power to detect

TABLE 1
Comparison of *Ddc* haplotype frequencies
between *SD* and *SD*⁺ chromosomes

Chromosome type	Haplo-A	Haplo-B	Haplo-C	Haplo-D	Total
<i>Segregation</i>					
<i>Distorter</i>	4	15	0	0	19
Wild type	22	11	10	11	54
Total	26	26	10	11	73

$P = 0.000015 \pm 0.0000039$ (SE due to binomial sampling).

linkage disequilibrium between *Ddc* and *SD* was maximized. The observed haplotype frequencies, based on the two restriction sites surveyed, are presented in Table 1. In this case, the data were not collapsed into two allelic classes; instead, a test of independence was conducted by a standard Monte Carlo permutation procedure (e.g., see ROFF and BENTZEN 1989), resulting in the value $P = 0.000015$. The accuracy of the computer program used for this test was assessed by comparison to Fisher's exact test across a series of 2×2 tables; in all cases, the simulation results were extremely close to the actual probabilities (data not shown). In the case of *Ddc*, if the data are instead collapsed into a 2×2 table (with resulting cell values of 43, 11, 4, and 15), the probability according to Fisher's exact test is $P = 0.0000092$, which is even lower than that observed by simulation with all four allelic classes. We concluded that *SD* and *Ddc* are in linkage disequilibrium. Furthermore, because the RFLP patterns of all *SD* chromosomes were identical to those for *SD*⁺ chromosomes (Table 1, blots not shown), there was no evidence of divergence between *SD* and *SD*⁺ at *Ddc*.

***Gapdh-1* near the base of 2R:** Finally, we surveyed 16 *SD* and 22 *SD*⁺ chromosomes for SSCPs at *Gapdh-1*, which is near the base of 2R and approximately 3 cM away from *Sd* and *Top2* (refer to Figure 1; LINDSLEY and ZIMM 1992). *Gapdh-1* was chosen because it is the most proximal euchromatic gene on 2R that had already been surveyed for polymorphism and divergence in the *D. melanogaster* species group (J. McDONALD and M. KREITMAN, unpublished results). Specifically, 512 bp from the 3' end of this gene were shown to exhibit fairly typical levels of polymorphism and divergence for these species, with two common haplotypes differentiated by two tightly linked base pair substitutions. We used the same polymerase chain reaction (PCR) primers these authors had used to amplify this 512-bp fragment from both *SD*⁺ and *SD* chromosomes and then used the SSCP technique (ORITA *et al.* 1989) to evaluate levels of polymorphism. The results were as follows: 16 *SD* chromosomes exhibited haplotype A, whereas three had haplotype B; in contrast, 10 *SD*⁺ chromosomes exhibited haplotype A, and 12 had haplotype B. The probability of observing this result by chance alone, under the assumption of linkage equilibrium, is $P = 0.021$,

TABLE 2
Comparison of *Gapdh-1* haplotype frequencies
between *SD* and *SD*⁺ chromosomes

Chromosome type	Haplotype A	Not Haplotype A	Total
<i>Segregation Distorter</i>	16	3	19
Wild type	10	12	22
Total	26	15	41

$P = 0.02151 \pm 0.00015$ (SE due to binomial sampling).

suggesting that the region in linkage disequilibrium with *SD* extends beyond the centromere to *Gapdh-1* (Table 2). Furthermore, none of the SSCP results gave any indication that the *SD* haplotypes at *Gapdh-1* had diverged from those found on *SD*⁺ chromosomes.

DISCUSSION

To investigate the evolutionary history of the *SD* system of meiotic drive, we surveyed both *SD* and *SD*⁺ chromosomes for levels of polymorphism and divergence at six loci that together encompass the second chromosome centromeric region (refer to Figure 1). Discussion of these results is organized according to the type of insight each gene provided into the evolutionary history of the *SD* complex. First, we consider the data that were most informative for each of the three major component loci separately. Next, analyses of linked euchromatic regions are discussed.

***Sd* evolved recently:** Our estimates of neutral polymorphism and divergence at the *Top2* first intron agreed closely with previous results for noncoding sequences of other euchromatic loci in these species (AQUADRO and BEGUN 1993; HUDSON *et al.* 1994; KREITMAN and WAYNE 1994). This suggests that the *Top2* intron is evolving in a manner similar to the introns of other euchromatic loci for these species. This result rules out evolutionary scenarios that invoke strong selection or unusual mutation pressures affecting the intron sequences directly.

If the haplotypes at *Top2* were in complete linkage equilibrium with the allelic state of *SD*, then it would be reasonable to expect a sample of *SD* chromosomes to provide an estimate of nucleotide diversity at *Top2* that agreed with an estimate based solely on *SD*⁺ chromosomes. One clear result of our analysis, however, is that alleles at *Top2* and *SD* are not randomly associated (*i.e.*, they are in linkage disequilibrium). Hence, it appears that a particular lineage of *Top2* haplotypes has been sequestered on *SD* chromosomes for an undetermined period of time. The amount of polymorphism among these haplotypes, as well as the divergence between sequences from *SD* *vs.* *SD*⁺ chromosomes, should be proportional to the time since an ancestral *Top2* sequence was sequestered on an *SD* chromosome.

If this sequestration happened when *SD* first evolved, then neutral divergence at *Top2* can be considered an

accurate measure of the age of the *SD* system. If there has instead been recombination events exchanging *Top2* haplotypes between *SD* and *SD*⁺ chromosomes, then this divergence will be an underestimate of the age of the system. Recombination of this sort seems extremely unlikely, however, given the position of *Top2* with respect to the components of the *SD* complex: First, *Top2* is situated right next to the *Sd* locus [one of the sequences sampled from the *SD-Mad* chromosome ends less than 20 base pairs from the proximal end of the tandem duplication that has been shown to be the *Sd* locus; POWERS and GANETZKY (1991); MCLEAN *et al.* (1994); PALOPOLI *et al.* (1994); PALOPOLI (1995)]; this location makes it unlikely that a recombination event will separate *Sd* and *Top2*. Second, *Top2* is located *between* the driving locus (*Sd*) and its target locus (*Rsp*), so that the rare single recombinants in this very small region are doomed to rapid extinction because the result is either a self-distorting, "suicide" chromosome, or a nondriving chromosome hampered by a deleterious allele at *Rsp*. Only double-recombination events—with one exchange between *Sd* and *Top2* and the other between *Top2* and *Rsp*—would be likely to persist in the population. Hence, we believe that *Top2* is likely to have been sequestered just once—when *SD* first evolved—and that the estimate of neutral divergence at *Top2* can be considered a true estimate of the age of the *SD* system.

Interestingly, there was no polymorphism among *Top2* sequences sampled from *SD* chromosomes and no divergence between *SD* and *SD*⁺ chromosomes (Figure 2). Furthermore, the sequence from one wild-type chromosome was identical to all five *SD* sequences for this intron. Based on this high degree of similarity between *SD* and *SD*⁺ sequences at *Top2*, we concluded that *SD* chromosomes are of quite recent origin; indeed, this result suggests that the *SD* complex is no older than the average coalescence time between pairs of randomly chosen *SD*⁺ alleles. Based on the complete linkage disequilibrium between *SD* and *Top2*, as well as the complete monomorphism among *SD* chromosomes, we further concluded that the *SD* chromosomes have themselves shared a common ancestor in the very recent past. The time since this recent common ancestor among *SD* chromosomes could either be the time since the origin of the *SD* system or the time since a secondary sweep among the *SD* chromosomes took place.

***SD-Rsp*⁺ evolved recently and consists of small blocks of repeats:** We also surveyed RFLPs and DNA sequences at the *Rsp* locus. Like many families of tandem-repetitive DNA, *Rsp* exhibits substantial RFLP variation among chromosomes (Figure 3; WU *et al.* 1988; WU and HAMMER 1991). In contrast to the high degree of variability exhibited by *SD*⁺ chromosomes, all *SD* chromosomes that were examined exhibited one of two similar RFLP patterns (Figure 4).

Based on the extensive data available for *SD*⁺ chromo-

somes, *Rsp* repeats appear to diverge extremely rapidly in DNA sequence (APPENDIX B; for additional details, see WU *et al.* 1988; WU and HAMMER 1991; CABOT *et al.* 1993). Given this high rate of nucleotide substitution, we can ask whether the *Rsp* repeats sequenced from *SD-Mad* have been sequestered on *SD* chromosomes for a long enough time to have diverged appreciably from *Rsp* repeats on wild-type chromosomes. The answer appears to be no—the average divergence between *SD* and *SD*⁺ repeats appears to be approximately 5%, which agrees well with the level of polymorphism among *SD*⁺ repeats. Again, consistent with the results for *Top2* sequences and *Rsp* RFLPs discussed above, these results give us no reason to believe that the *SD* complex is any older than the average coalescence time between two randomly chosen *SD*⁺ alleles at the *Rsp* locus [for an accessible review of the coalescent process, see HUDSON (1990)].

Taken together, these data for the *Rsp* arrays on *SD vs. SD*⁺ chromosomes support the main conclusion drawn from the *Top2* results—that *SD* arose in the very recent past. In addition, these data suggest that the region hitchhiking with *SD* extends from *Top2* to beyond the second chromosome centromere. Some caution must be exercised when interpreting results from *Rsp*, however, because this locus is known to be subjected to strong selective forces related to the *SD* system (see discussion in CHARLESWORTH and HARTL 1978; WU *et al.* 1989); furthermore, there appears to be a small number of *Rsp* repeats in the centric heterochromatin of the third chromosome (R. CAZZI, C. CAGGESE and S. PIMPINELLI, unpublished results), which would tend to confound these analyses. For these reasons, RFLP patterns at the *lt* locus were also examined.

***E(SD)* evolved recently:** Results for *lt* are particularly interesting because this gene is embedded in the same stretch of β -heterochromatin as the *E(SD)* locus (BRITTNACHER and GANETZKY 1984). Since *E(SD)* is a weak segregation distorter in its own right (SHARP *et al.* 1985; TEMIN 1991), it is a formal possibility that *E(SD)* was the original distorting locus, with the *Sd* locus having evolved as a more recent addition to the *SD* system (see LYTTLE 1991, 1993); according to this hypothesis, the results for *lt* would be expected to exhibit more *SD-SD*⁺ divergence than was observed for the *Top2* region. The more widely accepted scenario is that the *Sd* locus evolved first, followed by both enhancers and suppressors of distortion (*e.g.*, CHARLESWORTH and HARTL 1978).

The extremely low rates of crossing over in the β -heterochromatin of *Drosophila* (ASHBURNER 1989 and references therein) makes the molecular population genetics of *lt* a reasonable gauge of the evolutionary history of *E(SD)*. Regarding the recombination rates in heterochromatin, WILLIAMS *et al.* (1989) provide an estimate of 10^{-4} events per *X* chromosome per generation; unfortunately, the majority of these events were within

the recombinational hot spot of the ribosomal RNA locus. To our knowledge, there has only been one documented spontaneous recombination event within the autosomal heterochromatin of *Drosophila*: In a study of induced recombination in this region, TATTERSALL (1981; referenced in SINCLAIR *et al.* 1983) observed a single spontaneous recombination event between *lt* and *rolled* (the latter is located in the heterochromatin at the base of 2R).

The fact that *SD* chromosomes harbored the *most common* wild-type allele at *lt*, despite the fact that this locus is in the same stretch of β -heterochromatin as the *E(SD)* locus, lends strong support to the conclusion that *SD* evolved recently. In other words, there was once again no appreciable divergence between *SD* and *SD*⁺ alleles, giving us no reason to place the age of *SD* beyond the average time of coalescence among *SD*⁺ alleles. These results for the *lt* locus provide confirmation of the conclusions already drawn based on the *Top2* sequences and the RFLP/sequence comparisons at the *Rsp* locus.

***SD* hitchhiking effect stops short of *Adh* on 2L:** Since *SD* and *Adh* are apparently in linkage equilibrium, it is clear that the region hitchhiking with *SD* does *not* extend as far as the *Adh* gene. This is not particularly surprising, since *Adh* is located approximately 3.9 cM away from the *Sd* locus (and even further from the rest of the *SD* complex; refer to Figure 1); furthermore, the *Adh* locus is known to be held in a polymorphic state by balancing selection (KREITMAN and HUDSON 1991), which would tend to speed the decay of any initial disequilibrium between *SD* and *Adh*.

***SD* hitchhiking effect extends at least as far as *Ddc* on 2L:** Since the region hitchhiking with *SD* does not extend as far as *Adh* in this distal direction, an interesting place to look for linkage disequilibrium is somewhere between *Adh* and the *Sd* locus. Analysis of the *Ddc* gene region RFLPs suggested that the stretch of chromosome being swept along with *SD* extends distally on 2L at least as far as the *Ddc* gene (but not as far as *Adh*). The *Ddc* gene is well into the euchromatin of 2L. This result suggests that the initial disequilibrium between *SD* and *Ddc* has not had enough time to be broken down completely by recombination, despite the estimated 0.1 cM between these loci [Note, however, that secondary “sweeps” among *SD* haplotypes are thought to occur and could also explain this result; see TEMIN and MARTHAS (1984) for empirical evidence of such a sweep among *SD* chromosomes.].

***SD* hitchhiking effect extends to *Gapdh-1* on 2R:** This gene is located in the euchromatin near the base of 2R (LINDSLEY and ZIMM 1992). The SSCP results for *Gapdh-1* suggest that the stretch of chromosome being swept along with *SD* extends at least as far as *Gapdh-1* on 2R. Altogether, the hitchhiking region extends from the euchromatin at the base of 2L, across the entire centromeric heterochromatin, and well into the euchromatin

at the base of 2R. Perhaps even more striking was the observation that the three *SD* chromosomes that exhibited the "wrong" haplotype at *Gapdh-1* were as follows: *SD-LA*, *SD-V017B*, and *SD-Roma*. Not only are these the only three *SD* chromosomes sampled from the Mediterranean region (2 from Spain and 1 from Italy), but they are also the three weakest *SD* chromosomes that have been recovered directly from nature (e.g., see TEMIN *et al.* 1991). It is intriguing that one of the linked enhancers of *SD*, designated *M(SD)*, has been mapped to the same approximate location (HIRAIZUMI *et al.* 1980). Hence, in agreement with genetic analyses for this meiotic drive system, these data suggest that strong *SD* chromosomes harbor a specific complement of enhancers on 2R that are not shared by weak *SD* chromosomes. Finally, these data support the WU and HAMMER (1991) hypothesis of a Mediterranean origin for *SD* (although it is worth noting that no attempt has been made to sample *SD* from Africa).

General conclusions and implications: Based on a survey of polymorphism and divergence at six loci spanning the second chromosome centromere, it appears that *SD* has evolved so recently that there is essentially no detectable divergence (besides the actual loci involved in the distortion phenotype) between *SD* and *SD*⁺ haplotypes. The *SD* complex, with all of its interacting components, appears to have evolved between "ticks" of all of the molecular clocks that have been assayed here (even the rapidly evolving *Rsp* sequences and *lt* locus RFLPs). Furthermore, it appears that the entire second chromosome pericentromeric region is hitchhiking with the *SD* complex; this constitutes at least 10% of the *D. melanogaster* genome based on mitotic chromosome maps (ASHBURNER 1989).

As can be predicted by population genetics theory (PROUT *et al.* 1973; CHARLESWORTH and HARTL 1978; CROW 1991), drive enhancers are found linked in coupling to *SD*, and drive suppressors are either unlinked or in repulsion. These models also explain the observation that various inversions have accumulated on different *SD* chromosome lineages, thereby tightening linkage among *Sd*, *E(SD)*, *Rsp*ⁱ, and at least two additional enhancers of drive on 2R. The results obtained here represent one of the rare cases in which the selective forces responsible for a hitchhiking event can be identified unambiguously—perhaps the only such case in a region of low recombination.

Analogous but strikingly different results have been obtained for two other well-studied systems of meiotic drive. Molecular surveys have determined that both the Sex-Ratio chromosomes of *D. pseudoobscura* and the *t*-haplotypes of *Mus* are essentially monomorphic within their host species. In contrast to *SD*, however, the extent of divergence between the meiotic-drive and wild-type chromosomes in these systems is much higher than the standing levels of intraspecific polymorphism among wild-type chromosomes. These results suggest that both

of these meiotic drive systems are quite ancient (WU and HAMMER 1991; BABCOCK and ANDERSON 1996). Interestingly, identical *t*-haplotypes are found across a range of mouse species, indicating that a *trans*-species sweep has occurred among *t*-haplotypes (HAMMER *et al.* 1989; HAMMER 1991; WU and HAMMER 1991).

Although we cannot completely rule out the possibility, it seems extremely unlikely that the molecular survey reported here has missed an ancient polymorphism "bubble" (e.g., as observed in the *Adh* slow-fast polymorphism; KREITMAN and HUDSON 1991) in the *Sd* region for the following reasons: (1) We expect there to be strong proximal linkage between *Sd* and the centromere due to epistatic selection and hence only *double* recombinants within a very small region would be likely to enjoy a long sojourn time. (2) Given that *E(SD)* and extremely insensitive *Rsp* alleles are critical components of the system and that both are located in regions of almost zero recombination, the lack of divergence at both *lt* and *Rsp* becomes difficult to explain by such a hypothesis. (3) The fact that the *Sd* haplotypes are almost monomorphic means that if the ancient polymorphism scenario is true, then one must also postulate a recent sweep among *SD* chromosomes. (4) Recent work has shown that the tandem duplication is indeed the *Sd* locus itself (POWERS and GANETZKY 1991; MCLEAN *et al.* 1994; PALOPOLI and WU 1994), and based on the published restriction maps of this duplication, the two halves appear to have very similar sequences, which is entirely consistent with the interpretations presented above.

There has been considerable recent interest in the observation that low levels of DNA sequence polymorphism are typically associated with regions of low recombination in the *Drosophila* genome (STEPHAN and LANGLEY 1989; BEGUN and AQUADRO 1991, 1992; BERRY *et al.* 1991; STEPHAN and MITCHELL 1992; AQUADRO and BEGUN 1993; LANGLEY *et al.* 1993). The favored mechanism to explain this has been positive Darwinian selection "sweeping" away the polymorphisms; in regions of low recombination, large segments are carried along with every advantageous allele that is fixed nearby (BEGUN and AQUADRO 1992). The most reasonable alternative is the background selection model (CHARLESWORTH *et al.* 1993), which posits a reduction in effective population size for these regions because each deleterious mutation is destined to drag a relatively large segment to eventual loss. Based on the results reported here, it would seem that an important subcategory of the first model should be stressed when accounting for the correlation between levels of polymorphism and recombination rates: meiotic drive systems, which are expected on theoretical grounds to arise more readily in regions of low recombination (CHARLESWORTH and HARTL 1978; CROW 1991), can sweep quickly to appreciable frequencies and carry large segments of the chromosome along for the ride. This would be particularly

true if the meiotic drive alleles often reach fixation, rather than being held in a polymorphic state due to counterbalancing selection as has apparently happened in the *SD* system. Any multilocus, epistatically selected gene complex might be expected to exhibit the same sort of effects, particularly in regions of reduced recombination.

This work was completed by M.F.P. in partial fulfillment of the requirements for the Ph.D. in the Department of Ecology and Evolution, University of Chicago. We thank NORMAN JOHNSON, MICHAEL WADE, BRIAN CHARLESWORTH, MARTIN KREITMAN, and DOUGLAS CRAWFORD for stimulating discussions and criticisms of an earlier version of this manuscript. We thank BARBARA WAKIMOTO for sending us several clones of the *lt* gene as well as a deficiency stock for this region. We thank WINIFRED W. DOANE for sending us a genomic clone from the *Ddc* gene region. We also thank JOHN McDONALD and MARTIN KREITMAN for sharing unpublished *Gapdh-1* DNA sequence results, as well as PCR primers for this gene. This work was supported by a National Science Foundation (NSF) Graduate Fellowship and a National Institutes of Health (NIH) Genetics Training Grant to M.F.P., and both NSF and NIH grants to C.-I.W.

LITERATURE CITED

- AQUADRO, C. F., and D. J. BEGUN, 1993 Evidence for and implications of genetic hitchhiking in the *Drosophila* genome, pp. 159–178 in *Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology*, edited by N. TAKAHATA and A. G. CLARK. Japan Sci. Soc. Press, Tokyo.
- AQUADRO, C. F., R. M. JENNINGS, JR., M. M. BLAND, C. C. LAURIE and C. H. LANGLEY, 1992 Patterns of naturally occurring restriction map variation, Dopa decarboxylase activity variation and linkage disequilibrium in the *Ddc* gene region of *Drosophila melanogaster*. *Genetics* **132**: 443–452.
- ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- BABCOCK, C. S., and W. W. ANDERSON, 1996 Molecular evolution of the Sex-Ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. *Mol. Biol. Evol.* **13**: 297–308.
- BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow-achaete* region. *Genetics* **129**: 1147–1158.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**: 1111–1117.
- BRITTNACHER, J. G., and B. GANETZKY, 1983 On the components of segregation distortion in *Drosophila melanogaster*. II. Deletion mapping and dosage analysis of the *SD* locus. *Genetics* **103**: 659–673.
- BRITTNACHER, J. G., and B. GANETZKY, 1984 On the components of segregation distortion in *Drosophila melanogaster*. III. Nature of Enhancer of *SD*. *Genetics* **107**: 423–434.
- CABOT, E. L., P. DOSHI, M.-L. WU and C.-I. WU, 1993 Population genetics of tandem repeats in centromeric heterochromatin: unequal crossing over and chromosomal divergence at the *Responder* locus of *Drosophila melanogaster*. *Genetics* **135**: 477–487.
- CHARLESWORTH, B., and D. L. HARTL, 1978 Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89**: 171–192.
- CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CROW, J. F., 1991 Why is Mendelian segregation so exact? *BioEssays* **13**: 305–312.
- DELARBRE, C., Y. KASHI, P. BOURSOT, J. S. BECKMANN, P. KOURILSKY *et al.*, 1988 Phylogenetic distribution in the genus *Mus* of *t*-complex-specific DNA and protein markers: inferences on the origin of *t*-haplotypes. *Mol. Biol. Evol.* **5**: 120–133.
- DEVLIN, R. H., B. BINGHAM and B. T. WAKIMOTO, 1990 The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. *Genetics* **125**: 129–140.
- DOSHI, P., S. KAUSHAL, C. BENYAJATI and C.-I. WU, 1991 Molecular analysis of the Responder satellite DNA in *Drosophila melanogaster*: DNA bending, nucleosome structure, and Rsp-binding proteins. *Mol. Biol. Evol.* **8**: 721–741.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- FU, Y. X., and W. H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GANETZKY, B., 1977 On the components of segregation distortion in *Drosophila melanogaster*. *Genetics* **86**: 321–355.
- HALDANE, J. B. S., 1932 *The Causes of Evolution*. Princeton University Press, 1990 edition, Princeton, NJ.
- HAMMER, M. F., 1991 Molecular and chromosomal studies on the origin of *t* haplotypes in mice. *Am. Nat.* **137**: 359–365.
- HAMMER, M. F., J. SCHIMENTI and L. M. SILVER, 1989 Evolution of mouse chromosome 17 and the origin of inversions associated with *t* haplotypes. *Proc. Natl. Acad. Sci. USA* **86**: 3261–3265.
- HARTL, D. L., 1974 Genetic dissection of segregation distortion. I. Suicide combinations of *SD* genes. *Genetics* **76**: 477–486.
- HARTL, D. L., and A. G. CLARK, 1989 *Principles of Population Genetics*, Sinauer Associates, Sunderland, MA.
- HARTL, D. L., and Y. HIRAIZUMI, 1976 Segregation distortion, pp. 615–666 in *Genetics and Biology of Drosophila*, Vol. 1b, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.
- HARTL, D. L., Y. HIRAIZUMI and J. F. CROW, 1967 Evidence for sperm dysfunction as the mechanism of segregation distortion in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **58**: 2240–2245.
- HAUSCHTECK-JUNGEN, E., and D. L. HARTL, 1982 Defective histone transition during spermiogenesis in heterozygous *Segregation Distorter* males of *Drosophila melanogaster*. *Genetics* **101**: 57–69.
- HEBERT, P. D. N., and M. J. BEATON, 1989 *A Practical Handbook of Cellulose Acetate Electrophoresis*. Unpublished manuscript.
- HIRAIZUMI, Y., D. W. MARTIN and I. A. ECKSTRAND, 1980 A modified model of Segregation Distortion in *Drosophila melanogaster*. *Genetics* **95**: 693–706.
- HIRAIZUMI, Y., and K. NAKAZIMA, 1965 *SD* in a natural population of *D. melanogaster* in Japan. *Dros. Inf. Serv.* **40**: 72.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Survey of Evolutionary Biology*, Vol. 7, edited by D. FUTUYMA and J. ANTONOVICS. Oxford Univ. Press, New York.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- KIMURA, M., 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**: 565–582.
- KREITMAN, M., and M. L. WAYNE, 1994 Organization of genetic variation at the molecular level: lessons from *Drosophila*, pp. 157–183 in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. STRED, G. P. WAGNER and R. DESALLE. Birkhauser Verlag, Basel, Switzerland.
- LANGLEY, C. H., J. MACDONALD, N. MIYASHITA and M. AGUADÉ, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **90**: 1800–1803.
- LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, New York.
- LYTTLE, T. W., 1991 Segregation distorters. *Annu. Rev. Genet.* **25**: 511–557.
- LYTTLE, T. W., 1993 Cheaters sometimes prosper: distortion of Mendelian segregation by meiotic drive. *Trends Genet.* **9**: 205–210.
- MCLEAN, J. R., C. J. MERRILL, P. A. POWERS and B. GANETZKY, 1994 Functional identification of the *Segregation distorter* locus of *Drosophila melanogaster* by germline transformation. *Genetics* **137**: 201–209.

- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NICOLETTI, B., G. TRIPPA and A. DEMARCO, 1967 Reduced fertility in *SD* males and its bearing on segregation distortion in *Drosophila melanogaster*. *Atti Acad. Naz. Lincei* **43**: 383–392.
- OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G. ANDERSON *et al.*, 1982 Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution* **36**: 86–96.
- ORITA, M., Y. SUZUKI, T. SEKIYA and K. HAYASHI, 1989 Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**: 874–879.
- PALOPOLI, M. F., 1995 Molecular and evolutionary genetics of the *Segregation Distorter* system of meiotic drive in *Drosophila melanogaster*. Ph.D. thesis, University of Chicago.
- PALOPOLI, M. F., P. DOSHI, and C.-I. WU, 1994 Characterization of two *Segregation distorter* revertants: evidence that the tandem duplication is necessary for *Sd* activity in *Drosophila melanogaster*. *Genetics* **136**: 209–215.
- PILDER, S. H., M. F. HAMMER and L. M. SILVER, 1991 Novel mouse chromosome 17 hybrid sterility locus: implications for the origin of *t* haplotypes. *Genetics* **129**: 237–246.
- POWERS, P. A., and B. GANETZKY, 1991 On the components of segregation distortion in *Drosophila melanogaster*. V. Molecular analysis of the *Sd* locus. *Genetics* **129**: 133–144.
- PROUT, T., J. BUNDEGAARD and S. BRYANT, 1973 Population genetics of modifiers of meiotic drive. I. The solution of a special case and some general implications. *Theor. Pop. Biol.* **4**: 444–465.
- ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol. Biol. Evol.* **6**: 539–545.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SANDLER, L., and K. GOLIC, 1985 Segregation distortion in *Drosophila*. *Trends Genet.* **1**: 181–185.
- SHARP, C. B., A. J. HILLIKER and D. G. HOLM, 1985 Further characterization of genetic elements associated with the Segregation Distorter phenomenon in *Drosophila melanogaster*. *Genetics* **110**: 671–688.
- SINCLAIR, D. A., R. C. MOTTUS and T. A. GRIGLIATTI, 1983 Genes which suppress position-effect variegation in *Drosophila melanogaster* are clustered. *Mol. Gen. Genet.* **191**: 326–333.
- STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermillion* and *forked* loci. *Genetics* **121**: 89–99.
- STEPHAN, W., and S. J. MITCHELL, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. *Genetics* **132**: 1039–1045.
- TEMIN, R. G., 1991 The independent distorting ability of the *Enhancer of Segregation distortion*, *E(SD)*, in *Drosophila melanogaster*. *Genetics* **128**: 339–356.
- TEMIN, R. G., and M. MARTHAS, 1984 Factors influencing the effect of segregation distortion in natural populations of *Drosophila melanogaster*. *Genetics* **107**: 375–393.
- TEMIN, R. G., B. GANETZKY, P. A. POWERS, T. W. LYTLE, S. PIMPINELLI *et al.*, 1991 Segregation distortion in *Drosophila melanogaster*: genetic and molecular analyses. *Am. Nat.* **137**: 287–331.
- TOKUYASU, K. T., W. J. PEACOCK and R. W. HARDY, 1977 Dynamics of spermiogenesis in *Drosophila melanogaster*. VII. Effects of Segregation Distorter (*SD*) chromosome. *J. Ultrastruct. Res.* **58**: 96–107.
- WAKIMOTO, B. T., and M. G. HEARN, 1990 The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. *Genetics* **125**: 141–154.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WILLIAMS, S. M., J. A. KENNISON, L. G. ROBBINS, and C. STROBECK, 1989 Reciprocal recombination and the evolution of the ribosomal gene family of *Drosophila melanogaster*. *Genetics* **122**: 617–624.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- WU, C.-I., and M. F. HAMMER, 1991 Molecular evolution of ultraselfish genes of meiotic drive systems, pp. 177–203 in *Evolution at the Molecular Level*, edited by R. K. SELANDER, A. G. CLARK and T. WHITTAM. Sinauer Press, Sunderland, MA.
- WU, C.-I., T. W. LYTLE, M.-L. WU and G.-F. LIN, 1988 Association between a satellite DNA sequence and the *Responder* of Segregation Distorter in *D. melanogaster*. *Cell* **54**: 179–189.
- WU, C.-I., J. R. TRUE and N. JOHNSON, 1989 Fitness reduction associated with the deletion of a satellite DNA array. *Nature* **341**: 248–251.
- WYCKOFF, E., D. NATALIE, J. M. NOLAN, M. LEE and T.-S. HSIEH, 1989 Structure of the *Drosophila* DNA Topoisomerase II gene: nucleotide sequence and homology among topoisomerases II. *J. Mol. Biol.* **205**: 1–13.

Communicating editor: G. B. GOLDING

APPENDIX A: *Top2* FIRST INTRON SEQUENCES

[illegible]

APPENDIX B

Rsp sequences from *SD-Mad* (in shaded boxes) are aligned with wild-type repeats: “Left” repeats are above the dotted line and “Right” repeats are below. Double slash represents the sequence covered by primers and hence is not informative.

```

cb3-1 .....A.....C. // * .....
cb6-1 ..... // .....
cb9-1 ..... // ..T.....

lp13-1 ..... // .....
lp17-1 ..... // C.....
lp72-1 .....A.....T..... // .....

cs1-1 ...G.....A ... ..C. .... ..T. // .....
cs3-1 .....C.....G..... ..T. // .....
cs5-1 .....G..... ..T. // .....

```

```

Sd(B)1 .....C. .... ..T. // .....
Sd(C)1 ..... // .....
Sd(D)1 ...G.....A ... ..C. .... ..T. // .....
Sd(E)1 ..... ..T. // .....A.

```

```

cb3-2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // ...G.....
cb6-2 ....C....A. . .A. ..C... ..C A.....G.A .....A..... ..A.GT. ..T. // TT.....C....
cb9-2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A..T. . .

```

```

lp13-2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....C....
lp17-2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....C....
lp72-2 ....A.....A. .C.A. ..C...C..C A..T.G...A .....A..... // .....C....

```

```

cs3-2 ....C....A. . .A. ..C.A. ..C A.....A .....A..... ..A.GT. ..T. // ..A.....C....
cs5-2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....C....

```

```

Sd(A) CC....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....
Sd(B)2 ..... // ...G.....
Sd(C)2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....C....
Sd(D)2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....
Sd(E)2 ....C....A. . .A. ..C... ..C A.....A .....A..... ..A.GT. ..T. // .....C....

```

L-Con TCTAGAGATTCTGT TCA AC TGG TAAG CAAAAACAG TAAATTGCCTAAG TTTTA CA TTATA A // AACAGAAGATACCT