TESTS OF POPULATION GENETIC MODELS OF THE SEGREGATION DISTORTER SYSTEM IN WILD POPULATIONS OF *DROSOPHILA MELANOGASTER*

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOMEDICAL SCIENCES (GENETICS)

MAY 1996

By

John Bruce Anderson

Dissertation Committee:

Terrence Lyttle, Chairperson
Hampton Carson
David Haymer
John Hunt
Bruce Tabashnik

ABSTRACT

The Segregation Distorter system of meiotic drive in *Drosophila melanogaster* consists of a haplotype of second chromosome loci (*SD*) which together mediate the dysfunction of sperm carrying a variable number of 240 base pair satellite DNA repeats located in the centromere region of SD^+ homologs. These alleles are found worldwide in *D. melanogaster* populations in an apparently stable polymorphism that is unexpected under the simplest model of meiotic drive dynamics. The present study investigates whether a high level of repeat number mutation from insensitive (Rsp^i) to sensitive (Rsp^s) repeat number is the force that balances the drive effect of *SD* in removing Rsp^i alleles.

Iterative computer simulations of six hypothetical models of mutational processes produced equilibria suggesting that mutation-drive balance can produce stable polymorphism, but at frequencies different from those found in nature. Rates of mutation required to maintain polymorphism are on the order of 10^{-3} to 10^{-1} changes per generation.

Southern blot analysis of native forms of *SD* and *SD+* chromosomes from a natural Hawaii population indicated >100 and 18 copies of the *Rsp* repeat in canonical Rsp^s and Rsp^i forms respectively. This agrees with previous studies showing the correlation between copy number and sensitivity. No changes in repeat copy number in chromosomal isolines were observed through approximately 25 generations.

The three native isolines were used to establish population cages, which were sampled at intervals for *Sd*, Rsp^i *and* Rsp^s frequencies. After 300 days, Rsp^s was lost from all cages, which therefore did not reproduce the polymorphism observed in nature. Minimum χ^2 analysis shows that none of the six mutational models was a good predictor of the observed frequencies. Therefore, the hypothesized models are not supported by these data.

Cages containing only native Rsp^i and Rsp^s were established and changes in *Rsp* frequencies measured. Estimates of the relative fitness of the alleles the three cages was s = 0.087, 0.025 and 0.068 respectively. These selective coefficients between *Rsp* alleles from the same natural population are much less than those found in

previous studies using lab chromosomes and are too low to explain the observed *SD* polymorphism in this population.

Acknowledgments

I would like to thank Dr. Lyttle for making this research possible and for his constant intellectual stimulation. I would like to thank Kathy, Joan, Leyla, Sue, Dave, Mike, Buddy and Jayma for all the aid in the lab and for the friendship and support they showed.

I feel a great debt to all those who never stopped believing in in me over the years, especially my parents and all the others who love me. May God care for you as you cared for me.

And a special thanks, all this and more, to Roxanne, who was with me all the way. I wouldn't have made it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

1.1.1 Mendelian Segregation and Meiotic Drive

One of the fundamental rules of the science of genetics is Mendel's Law of Segregation. It is a constant but often unstated assumption in all studies of heredity and population genetics that heterozygotes will transmit their two alleles equally to their gametes.

Mendelian segregation is, in a sense, an altruistic system since an allele or chromosome allows its homolog an equal chance to be passed to the next generation. Such a system should be susceptible to invasion by selfish alleles that prevent the transmission of their homologs and thereby increase their own relative frequency of transmission (Crow, 1979, 1988, 1991; Eshel, 1985; Hurst and Pomiankowski, 1991; Liberman, 1976; Prout et al, 1973). This phenomenon is known as meiotic drive. Alleles that distort normal meiosis to increase their own relative transmission above Mendelian proportions are called *meiotic drive* alleles or *segregation distorters*. James Crow (1988) used the term "ultraselfish" to describe such genes, since they act not just for their own advantage, but actually damage their homologs.

There are several well-studied instances of meiotic drive (extensively reviewed in Lyttle, 1991; 1993). These include *SD* in *Drosophila*, *t*-haplotypes in *Mus*, *SR* in *Drosophila*, *MD* in *Aedes* and *Sk* in *Neurospora*. The molecular mechanism of meiotic drive is not well understood for any case but these systems share certain features.

1.1.2. Common Features of Meiotic Drive Systems

Meiotic drive systems are usually multilocus systems, involving at minimum a drive locus and a target locus. The drive is trans-acting (except for *MD*) and the target is cis-acting. A drive haplotype consists of a driving allele tightly linked to an insensitive target allele. The target haplotype consists of a non-driving allele linked to a sensitive target. Insensitive haplotypes consist of a non-driving allele linked to an insensitive target. Often, there are other closely linked modifiers. At the extreme, the tight linkage of several elements results in a drive haplotype in which a large portion of a chromosome is involved, as in the t-haplotypes of *Mus musculus* and *M. domesticus* in which approximately one-third of chromosome 17 is involved in the haplotype (Silver, 1985).

Meiotic drive is usually sex specific, with drive in males being most common. In most cases the drive is generated premeiotically by the trans-acting elements but takes the form of postmeiotic gametic competition, with a certain proportion of the target-bearing gametes

being disabled. It is likely that this type of drive usually takes place in the males for the following reason. If there is a proportional drop in fecundity due to the action of the drive allele, then there is no net drive since it will not be present in any more offspring than it would have been without drive. Because males produce an excess of sperm, a loss of up to half the sperm may not result in loss of fecundity. This is typical of *genic drive* as opposed to *chromosomal drive* which occurs preferentially in females (Lyttle, 1993).

Drive haplotype homozygotes are often infertile or sterile in the driving sex (Hartl, 1973). This is not the direct result of drive since the target locus linked to the driver is usually insensitive to the action of drive. It is more likely that these are the result of deleterious recessives in linkage with the drive locus. These develop via a process like Muller's Ratchet (Muller, 1964; Felsenstein, 1974) and tend to accumulate due to the reduced recombination and meiotic advantage of the drive chromosomes (Lyttle, 1993).

The general phenomenon of meiotic drive is of interest at two levels. Since these ultraselfish genes subvert Mendelian segregation, they can provide important insight into how normal segregation works at the cytological and molecular levels. Second, understanding the mechanisms by which an allele can cheat will enhance understanding of the point in meiosis that is disrupted.

Segregation distortion is also very interesting from a population genetics point of view since the distortion gives such a strong evolutionary advantage to the distorting allele. That is the

primary focus of this study. Of specific interest are the evolutionary dynamics that may act to prevent alleles that cheat Mendelian segregation from dominating a system. There have been numerous attempts, both theoretical and experimental, to understand the population dynamics of segregation distortion systems. These studies have provided important insight into how such systems should behave given a particular model and sets of parameters.

An allele that is able to increase its own relative fitness by reducing the transmission of its counterpart (i.e., distort segregation) should increase its frequency in the population rapidly to fixation. At that point, it would appear that there is no distortion, if the distorting allele does not work against copies of itself, or at least become invisible (Charlesworth and Hartl, 1978). This may be the reason that segregation distortion appears to be uncommon, but this may not be very likely since (1) that would create a genome overloaded with largely non-functional genes and (2) the mechanisms necessary to affect segregation ratio could have deleterious effects on fitness if there were many such genes in the genome.

An alternate pathway to stability is the development of insensitive targets. If there are insensitives present, the selective force of the drive should cause sensitives to decrease in frequency to zero. At this point, the frequency of the drive haplotype would be frozen since all targets are insensitive and drive becomes inoperable.

However, in most known cases of drive, there is a stable polymorphism of sensitive and insensitive alleles.

The fact that there are several known strong segregation distorters and targets that exist at intermediate frequencies in natural populations demonstrates that polymorphisms can be maintained but the fact that they are not more common suggests that there is some sort of evolutionary constraint on their proliferation. Eshel (1984) has argued that the fact that many of these polymorphisms are apparently stable over long periods suggests that there may be some general phenomenon that acts similarly in several systems to limit their frequencies. Simple mechanistic models are inadequate to explain this observation. The answer may lie in complex interactions at a level only accessible by population genetics.

1.1.3 Models of Meiotic Drive

A number of theoretical population genetics models of meiotic drive have been published (e.g.. Charlesworth and Hartl, 1978; Eshel, 1985; Eshel and Feldman, 1982; Hartl 1972; Lieberman, 1976; Liberman and Feldman, 1982; Prout et al 1973; Thomson and Feldman, 1974; 1976). These may be placed into two main classes that differ in their causative agents and therefore differ significantly in their interpretation and evolutionary implications for the SD system. The first, referred to here as selection models, are those that depend on fitness differentials among the components of the system.

The second is a broader grouping that includes models that look for neutral (non-selective) explanations by considering complex interactions produced by more detailed modeling of the mechanisms of the system. A consideration and comparison of these two classes of model represents a major part of this thesis.

1.1.3.1 Selection Models

Most of the general models of meiotic drive (e.g.. Prout et al 1973; Charlesworth and Hartl, 1978) depend on selection against the driving allele and against the insensitive allele for stable equilibrium. The model that has been most widely accepted is the classic one produced by Charlesworth and Hartl (1978) which depends on a drive-selection balance to maintain polymorphism of the drive components. The advantage of segregation distortion is countered by reduced fertility of the drive homozygotes and the tendency of insensitives to increase in frequency is countered by reduced fitness of the insensitives relative to sensitives. This result does not explain how several very different natural systems could be controlled since at least some of them are distinct in their mechanisms and therefore presumably subjected to different selective regimes. There is no *a priori* reason to suppose that an allele that distorts segregation must necessarily be deleterious in ways not directly associated with drive itself.

When the predictions of such theoretical models have been tested in experimental lab studies (e.g., Hiraizumi et al, 1960; Hartl, 1977), results have only partially supported the models. However, none have yet succeeded in predicting or explaining what is actually observed in natural populations. For example, *t*-haplotypes have been shown to have complete sterility and a high proportion of embryonic death in homozygotes (Klein, et al. 1984). The polymorphism is presumably maintained by the balance between this diploid selection and the haploid advantage of drive. However, the actual measured fitness difference and drive rates predict a higher frequency of *t* -haplotypes than the 10-20% observed in nature. The disagreement of the observation with the prediction suggests that the models are not completely accurate representations of the natural system. Likely explanations in the case of *t* include the effects of population structure and mating behavior (Lennington et al., 1988; Lennington and Egid, 1989).

1.1.3.2 Neutral Models

The other models to be considered are those in which the drive and target alleles are selectively neutral, except for the effects of drive itself. The polymorphism of drive alleles is maintained by interaction of the drive alleles with other loci or by intricacies of the meiotic drive process itself. Fitness effects occur at the level of long

term transmission advantage, rather than by direct selection at the allelic level.

There are fewer examples of this type of model. One group emphasizes the importance of the evolution of genetic suppressors in limiting the spread of driving alleles (e.g.. Stalker, 1961; Feldman and Otto, 1991; Bengtsson and Uyenoyama, 1990; Haig and Grafen, 1991; Liberman, 1990; Dominguez, et al, 1993). In such a model, the demographic effects of meiotic drive produce strong selection on the organism, at the diploid level, to develop mechanisms to suppress the effects of drive. A key result of these models is that a locus that suppresses drive generally need not be linked to the distorter locus to work. On the other hand, it has been repeatedly demonstrated that, for a two locus system with a killer and a target, drive can arise and persist only if the loci are tightly linked (e.g.. Haig and Grafen, 1991; Hurst and Pomiankawski, 1991; Lessard, 1985; Prout et al, 1973). Therefore, an increase in recombination itself can be an evolutionary response to counter meiotic drive (Haig and Grafen, 1991). These observations have important implications for the evolution of recombination but are outside the scope of this study.

The type of mechanistic model of meiotic drive in which we are interested is one in which polymorphism is maintained by interaction of the drive alleles themselves. It depends on closer examination of the alleles and their interaction as more than simply killer and target. This type of model is therefore dependent on special characteristics of a specific system. While this makes the

model less general, it may be that meiotic drive systems are relatively rare precisely because only systems that have such intrinsic controlling mechanisms can persist.

In this study, we consider such a model in which inherent characteristics of the system make it possible for a polymorphism to persist. The model involves a balance between drive and mutation at the target locus. The specific case is the *Segregation Distorter* system in *Drosophila melanogaster.*

1.1.4 The Segregation Distorter System

One of the best characterized systems of meiotic drive is the Segregation Distorter (*SD*) system of *Drosophila melanogaster. SD* was first discovered in 1956 (Hiraizumi and Crow, 1960). It was found that a male heterozygous for an *SD* chromosome and a sensitive *SD+* homolog produces a very high proportion (typically 90 to 100%) *SD* bearing sperm rather than the 50% expected under Mendelian segregation. The standard measure of drive strength, k, is defined as the proportion of *SD*-bearing sperm produced by an *SD/SD+* heterozygous male.

In the intervening years, this system has been extensively studied. Although the exact mechanism of the drive is still not known, a wealth of information on cytological, molecular and population aspects of the system has accumulated. This information has been the focus of some excellent recent reviews (Lyttle 1991;

1993; Temin et al, 1991, Wu and Hammer, 1990) and only details of pertinence to the present study are presented here.

The *SD* system consists of two main elements, Segregation Distorter (*Sd*) and Responder (*Rsp*), and several more or less well characterized modifiers that affect the distortion either positively or negatively. The drive locus (*Sd*) maps cytologically to 37D2-6 of 2L (Brittnacher and Ganetzky, 1983). The *Sd* allele gives a novel 12Kb restriction fragment containing a tandem duplication of a 5Kb DNA segment. Sd^+ (non-driving wildtype) gives a 7Kb fragment (McClean *et. al.*, 1994). This 12Kb fragment restores detectable levels of segregation distortion when transformed into Sd^+ germlines (McClean *et. al.*, 1994; Palopoli, *et. al.*, 1994). There is no known functional product produced by the wild-type allele (Sd^+) (Brittnacher and Ganetzky, 1983; Powers and Ganetzky, 1991), but several small transcripts have been identified. In addition to the Sd^+ transcripts, the *Sd* allele produces a unique 4 Kb transcript. The translation apparently involves splicing of several scattered exons and is not fully understood at this time (Ganetzky, pers. comm.), nor is the polypeptide that it may produce (McClean et al., 1994).

One major modifier, Enhancer of SD [*E(SD)*], which is closely linked to *Sd*, appears to be essential for distortion and may in fact be capable of distortion itself (Sharp et al, 1985; Temin, 1991). Other positive modifiers, such as *St(SD)* and *M(SD)*, also tend to be linked to the drive locus. *SD* chromosomes often have one or more inversions, such as the cosmopolitan *In(2R)NS* and a small pericentric inversion,

which have the effect of reducing recombination. For the purpose of the present study, the modifiers may be assumed to be present and stable and therefore, the drive complex may be treated as a single locus (*SD*).

The target locus (*Rsp*) is in the proximal heterochromatin of 2R. Several different alleles for *Rsp* have been isolated from nature. These range in sensitivity to *SD* from completely insensitive (Rsp^i) through standard sensitivity (Rsp^s) to supersensitive (Rsp^{ss}). In natural populations, 50-80% of the SD^+ chromosomes are found to be sensitive (Temin and Marthas, 1984). These alleles have been shown to be associated with a set of 240 bp, AT rich tandem repeats (Wu, et al, 1988). Sensitivity is directly proportional to the number of repeats, with Rsp^i having tens of copies, Rsp^s having hundreds and Rsp^{ss} having thousands (Wu and Hammer, 1990). The repeats are apparently arranged in stretches of tandem repeats, interspersed with unique, non-repeat sequences (Cabot et al., 1993). The alleles in nature carry high variability for repeat copy number and arrangement. In a sample of 60 second chromosomes from the same population that were restriction enzyme digested, Southern blotted and probed with a *Rsp* specific probe (producing DNA "fingerprints" for the *Rsp* locus in the same manner described herein), 59 distinct banding patterns were observed (Wu and Hammer, 1990).

The *Rsp* alleles have no apparent phenotype apart from sensitivity, although it has been shown that the *Rsp* repeat structure has protein binding properties (Doshi, et al. 1991). This, together

with its position near the centromere and its role in abnormal segregation, is consistent with it playing a role in chromatin structural regulation. However, deletion of the *Rsp*^{s} repeats leads to no obvious negative effect on spermatogenesis (Ganetzky, 1977), indicating that it has no essential wild-type function (Sandler and Golic, 1985). One study has presented data that suggests that *Rsp*^{s} has a slight fitness advantage over *Rsp*^{i} in cage population competitions (Wu et al., 1989). This suggests that there is some function that is advantageous but not required. However, this result is not entirely convincing. The comparison was between a sensitive chromosome and a chromosome derived from it by radiation-induced deletion of a relatively large region that included the repeat arrays. The reduction in fitness observed could have been due to loss of the intervening sequences or possibly to other, cryptic, radiation induced damage.

Research on the molecular and genetic characteristics of the loci of the Segregation Distorter system is being vigorously pursued in the laboratories of T. W. Lyttle and many others. Doubtless, new information will soon be forthcoming that will clarify many of the puzzling aspects of SD. For the purposes of the present study, the precise mechanisms of segregation distortion are of less immediate interest than how they influence the inheritance and transmission of the alleles in a population.

1.1.5 Population Dynamics of the Segregation Distorter System

SD-bearing chromosomes have been isolated from *D. melanogaster* populations from around the world. Virtually wherever *D. melanogaster* is found, *Sd* is in low but constant frequency of 1 to 5% (Lyttle, 1993). This observation is at the heart of the interest in *SD* among population geneticists because such a low but persistent frequency is completely at odds with predictions. There are two trends to consider in the dynamics of this system. First is the tendency of *SD* to increase its frequency and second is the tendency of Rsp^s to decrease in frequency. The failure of SD to increase to eliminate Rsp^s in natural populations has fueled theoretical and experimental studies for years.

One possible explanation is that the observed uniformity of frequencies could simply be an artifact. As commensals of man, these "garbage can" flies are regularly transported from one place to another and so the different populations in which we see *SD* may not be truly isolated. There are at minimum three good reasons to discount this explanation. First, there are at least two distinct cytological and molecular clades of the *SD* chromosomes (Trippa et al, 1980; Trippa, et al, 1981; Wu and Hammer, 1990). Secondly, both *SD* and *Rsp* alleles differ functionally, in terms of drive strength and cross-sensitivity, among populations (Trippa and Loverre, 1975; Shimikawa, 1987). These observations argue that it is not likely that

worldwide *SD* chromosomes are recently derived from the same population and maintained by high levels of migration. Third, there have been a host of studies indicating geographic differentiation among *D. melanogaster* populations for other loci. There is a great deal of chromosomal inversion polymorphism, with most natural populations harboring endemic inversions at low frequency and there is extensive literature on environmentally correlated clines in frequencies of cosmopolitan inversions (reviewed by Lemeunier and Aulard, 1992). There are also examples of geographic variation and latitudinal clines in allozyme polymorphisms (Oakeshott et. al., 1981; 1983, Lemeunier and Aulard, 1992). Clearly, populations of *D. melanogaster* are highly structured and a large amount of heterogeneity exists among populations.

A better explanation is that there are specific characteristics of *SD* action that produce a stable equilibrium frequency of the genotypes. There are two major classes of phenomena that could be responsible. First, there may be some selective advantage to specific genotypes that opposes the negative effects of drive. The second possibility is that an as yet undescribed characteristic of the system produces the observed dynamics. In particular, since sensitivity is proportional to the number of copies of a 240bp repeat, (Wu and Hammer, 1990) some process that causes a high rate of change in number of copies could maintain sensitivity in the face of pressure by meiotic drive against it.

Invoking selection against the insensitive allele or against the driving allele as the limiting factor, as do many models for *SD* in *D. melanogaster* (e.g.. Charlesworth and Hartl, 1978), is unsatisfying for a number of reasons. In particular, it is *ad hoc*, in that the degree and direction of the selection proposed is chosen to make the model fit the data, without good data for parameterization of the selection. Charlesworth and Hartl (1978) recognize this as a limitation of their model. Further, the model is not robust, in that relatively small changes in the values for selection make a large difference in the equilibrium value of the system. This weakens the model since a cosmopolitan species like *D. melanogaster* would be expected to experience different selective pressures in different parts of its range, producing different equilibrium frequencies. This is contrary to what is observed in nature. In fact, the frequencies are approximately the same in all the many environments, from nearctic to tropical, where *D. melanogaster* thrives. Lastly, there is a lack of convincing evidence that natural alleles of the components have a substantial selective differential (Lyttle, et. al. 1993; Wu et. al. 1988).

The other group of explanations for the maintenance stable population frequencies does not depend on outside forces, but rather interactions of processes inherent to the system. One type of model, originally suggested by Hickey et al (1986), is based on the unique structure of *Responder*. The observations that sensitivity exists as a continuum in nature rather than as discrete states and that there is high variability in the number of repeats suggests that mutability of

the locus may be important in the system. Repetitive sequences have a number of interesting properties that may suggest possible mechanisms for the action of the system. These mechanisms depend primarily on misalignment during replication, damage repair or crossover events and will be discussed in some detail in a later section.

Most of the classical experiments exploring the population dynamics of segregation distortion in *D. melanogaster* (e.g.. Hiraizumi et al, 1960; Hartl, 1977) have used long-established lines, which have had standard laboratory backgrounds introgressed. This approach is useful in that it isolates the gene of interest in a uniform background for testing. However, it raises a number of objections when the goal is to study the natural system. First of all, the experiments usually rely on phenotypic marker mutations (e.g.. *cn* and *bw*) on the chromosomes of interest which may introduce subtle fitness effects.

Secondly, the genetic backgrounds are uniform and may lack genes that are important to the function of the SD system in its natural state. These experiments have failed to reproduce the frequencies of the SD components that are found in nature. It may be that these artificial populations do not mimic the natural system in that some unknown components are missing or modified in the laboratory lines or that the experimental setup is not a realistic reproduction of the natural system.

Another problem is that the introgression of recently collected natural chromosomes into standard laboratory backgrounds raises

the possibility of the induction of hybrid dysgenesis. Natural populations are known to carry a number of transposable elements, such as P, that are suppressed in their native cytoplasmic background (Engels, 1989). When P-bearing strains are crossed to M cytotypes, which lack P elements and do not suppress transposition, there is an increase in transposition. This results in a host of related effects, grouped under the term hybrid dysgenesis, that are a direct result of the process of excision and insertion of the transposable element. These include mutation, chromosome rearrangement, male recombination, chromosome nondisjuction and sterility. Most wild populations have P-elements in almost all lineages (Anxolabehere, et al, 1985), while many laboratory strains are M cytotypes (Engels, 1989). Therefore, strains for crosses between wild-caught flies and laboratory flies must be carefully chosen and monitored for these effects.

Lastly, it also may be that the response of the particular chromosomes that have been taken as "standard" may in fact be unique and significantly different from that of chromosomes found in nature. It has been shown (Trippa and Loverre, 1975; Shimikawa, 1987) that the drive strengths of distorters against responders from different populations may not be the same as those of the same population. That is, a responder is typically more sensitive to its own distorter than to a foreign distorter but the opposite also occurs. This strongly suggests that the interacting components in a population have evolved in response to one another. The components should not

be thought of as fixed entities but are better considered as a constantly changing and finely tuned system. An experiment combining components from several different populations may not be the truest representation of a natural system.

An opportunity to explore these questions in a natural system in a new way is provided by the recent advances in understanding of the molecular basis of SD in *D. melanogaster*. Specifically, the availability of DNA probes that can identify both major SD component genotypes in otherwise wild-type flies makes it possible to examine population dynamics in flies from any source, independent of genetic markers. The present study develops several neutral models based on possible mechanisms affecting SD components and then tests the predictions of those models using cage populations derived from isolines of wild-caught *D. melanogaster*.

1.2 Experimental Approach

This research uses a two pronged approach: computer simulations and population cages using well-characterized wild chromosomes from nature. It is important to bear in mind that we are not specifically studying the molecular evolution of the sequences, but rather are interested in the population dynamics of existing alleles that can transmutate. Thus, we are following frequencies of alleles in the population rather than single

chromosome lineages. There are three different but related experiments.

The first set of experiments is an attempt to reproduce the phenomenon in the laboratory. A population cage provides optimal food, temperature and growth conditions, which should minimize the effects of selection. If the observed natural frequencies are reproduced in a population cage, it argues against environmentally dependent selective forces controlling the system. The second product of using well-characterized natural chromosomes in a population cage is that it allows observation of individual chromosome types under the action of segregation distorter. If interaction of the components produces physical changes is the alleles (specifically Responder), as has been hypothesized, there should be observable changes in the DNA fingerprints of the alleles.

Secondly, a set of experiments was designed to test the relative fitness of natural chromosomes using population cages. These experiments used the same chromosomes as the earlier set of experiments, except that the SD chromosome was excluded. In this way, changes in frequencies in the cages will presumably be due to fitness differences in the chromosomes.

Thirdly, computer simulation was used to make predictions about the system under various conditions. The computer simulations are a way to reproduce the system stripped of all variables except those whose importance we wish to test. Population genetics models of a system such as this are complex, involving

higher order equations and frequency dependent interactions. Such models are mathematically intractable, making an analytical approach unfeasible. An iterative simulation can produce the desired predictions about equilibria in a way that has the added advantage of producing data on predicted intermediate (approach to equilibrium) states of the system.

1.3 Models of Mutational Mechanisms

1.3.1 Background on Gross Structure of Highly Repetitive Tandem Repeats

For the purposes of this study, we are interested in possible mutational mechanisms that may tend to increase copy number in highly repetitive tandem repeat (HRTR) sequences. A great deal is known about these sequences and there is a wealth of literature on the dynamics of repeat number change.

Highly repetitive tandem repeats (HRTR) are relatively short, simple DNA sequences that are found both in protein coding regions of genes and in non-coding regions. HRTR's form a significant portion of heterochromatic and telomeric regions of chromosomes (Thompson-Stewart, et. al., 1994). The HRTR's considered here do not include amplified genes, such as rRNA's and chorion genes, which may have entirely different evolutionary causes and consequences

(Dover, 1982; Hancock and Dover, 1988; Ohta, 1981, Ohta, 1983). HRTR's can be divided into microsatellite (di- and tri-nucleotides) and minisatellite (10's or 100's of nucleotides per repeat). Minisatellite sequences typically occur in copy numbers of tens to hundreds while microsatellites have fewer copies (Harding et al, 1992). The differences between these two classes may have an impact on the mechanisms of copy number change and so the following discussion is conveniently divided into two parts, very short repeats (trinucleotides) and longer repeats.

1.3.1.1 Trinucleotide Repeat Variation

The current interest in the structure of highly repetitive DNA has been stimulated, at least in part, by the implication of hypervariability of repetitive DNA as the cause of several medically important genetic disorders. These are characterized by hypermutability of alleles and correlation between the size of the repeat region and severity of the condition (McMurray, 1995). The repeat unit in many cases is a trinucleotide. Examples include the following:

Huntington's Disease (MacDonald *et. al.*, 1993) is due to changes in the number of a trinucleotide repeat unit (CAG) that occurs in the coding region of IT15. IT15 is a 210 Kb segment of the candidate HD region that codes for a 348 Kd putative protein. The polymorphic trinucleotide repeat accounts for at least 17 "normal" alleles, varying from 11 to 34 CAG repeats. On HD chromosomes, the number of

repeats ranges from 42 to over 66 copies, with an apparent correlation with age of onset. The HD allele is dominant.

The mutation rate in HD is considered to be quite low. However, the Huntington's Disease Collaborative Research Group (MacDonald et al, 1993) examined a single chromosome through 3 generations in a lineage and found that the copy number went from 33 (normal, unaffected individual) to 49 (2 affected individuals). Another chromosome in the same family went from 36 to 44 in 3 generations. Given that low copy number chromosomes (<20 copies) are not known to increase dramatically, this suggests that high copy number may predispose expansion. While the mechanism of amplification has not been ascertained, it is presumed to involve slippage during replication, followed by misalignment due to the repeat structure. Repair of the resulting lesion results in expansion or contraction of the number of repeats. This mechanism, known as replication slippage repair (RSR), is described in detail in a subsequent section. Computer simulations of the populations genetics of this system show that simple length-dependent mutational bias towards longer alleles is sufficient to explain the incidence of HD (Rubinstein, et al, 1994).

Fragile-X syndrome is associated with the expansion of a CGG trinucleotide in the 5' untranslated region of the FMR1 gene (Fu et al, 1991), with greater than 52 units producing a phenotypic effect. The mutation rate is apparently quite high in this case, with up to one third of affected chromosomes newly arising in each generation (Sved and Laird, 1990).

Myotonic dystrophy is due to an unstable number of CTG repeats in the 3' untranslated region of the myotonin protein kinase gene (Aslanidis et al, 1992; Buxton et al, 1992). Normal individuals have 5 to 30 copies while affected individuals may have thousands (Harley et al, 1992). These alleles produce a dominant phenotype. In both these cases, hypermutability has been observed, with both increases and decreases in copy number between generations. Gennarelli et al. (1994) reported that the trinucleotide hypermutability that produces myotonic dystrophy is more pronounced in fathers than mothers.

Autosomal dominant dentatorubral-pallidoluysian atrophy (DRPLA) and Machado-Joseph disease (MJD) are neurodegenerative disorders caused by CAG trinucleotide repeat expansions (Ikeuchi, et al, 1995). These also show an inverse correlation of age of onset with the length of expanded repeats and greater instability in repeat lengths males than in females.

Spino-bulbar muscular atrophy has a CAG repeat in the coding region of an androgen receptor gene (LaSpada et al, 1991), but appears to be less mutable (Biancalana et al, 1992).

These examples suggest that trinucleotide variation that occurs in coding regions tolerates less mutability than that in non-coding regions. It is probably best to consider the two types separately because each has different constraints on variability. Trinucleotide variation inside an exon would produce a polypeptide with additional amino acids but no frameshift. A small number might lead to altered

or reduced function of the polypeptide while larger numbers would be expected to cause loss of function.

On the other hand, variation in numbers of trinucleotide in a non-coding region that produces a change in phenotype is presumably due to an effect on a genetic control structure or on a chromosomal scaffold structure. In this case, 3 dimensional structure is what is important and one might speculate that more "play" in the sequence might be tolerable before function is affected. Larger numbers of repeats would also presumably increase the likelihood of mispairing, which is an important feature in most models of copy number change.

The *Drosophila melanogaster* locus *Mastermind* produces several transcripts that are important in embryonic neurogenic development. One of the cDNA's contains a large and variable number of homopolymer peptide codons for glutamine, glycine and asparagine (Smoller et al, 1990) separating segments of unique sequence polypeptides. The protein was found to accumulate in the nucleus of early embryonic cells, indicating that it has a regulatory function.

Comparisons within *D. melanogaster* and between *D. melanogaster* and *D. virilis* (Newfeld et al, 1993) show that functional domains are conserved and that the homopolymers may serve as flexible spacers between them. Newfeld et al (1994) also suggest that the observed allelic variation in length follows a distribution that appears to be constrained by natural selection, with excessive

changes being deleterious. There is no evidence for which mechanism may account for mutability in this case.

It is interesting to note that many of the human genetic disorders associated with unstable repeat segments also display unusual segregation patterns that have some of the characteristics of meiotic drive systems (Klitz *et. al,* 1987). Gennarelli et al. (1994) conducted a study involving 897 sibs in 251 families affected with myotonic dystrophy. They found that the mutant alleles had preferential transmission, with 58.1% of offspring receiving the mutant allele and 41.9% receiving the normal allele. Additionally, this effect was more pronounced in fathers than in mothers with males producing 59.8% and females producing 55.6% affected offspring.

Autosomal dominant dentatorubral-pallidoluysian atrophy (DRPLA) and Machado-Joseph disease (MJD) also show meiotic drive (Ikeuchi, et al, 1995). In 211 transmissions in 24 DRPLA pedigrees and 80 transmissions in 7 MJD pedigrees, the mutant alleles were transmitted to 62.4% of offspring in DRPLA and 72.7% in MJD. As in the other cases of meiotic drive, both hypermutability and segregation distortion were more pronounced in males than in females.

These examples lack some of the usual features of meiotic drive systems. In particular, they are not systems which have a drive locus and a target locus. Rather, the increased number of repeats itself seems to produce the aberrant segregation. Even so, the

association of hypervariability of a repeat segment with an altered phenotype, the observation that the instability is more pronounced in males and the apparent non-Mendelian segregation of alleles all bear intriguing similarities to the Segregation Distorter system.

1.3.1.2 Longer Repetitive Tandem Repeats

Locus *D1Z2*, near the tip of the short arm of chromosome 1 in humans is implicated in Bloom's syndrome, a recessively transmitted genetic disorder (Groden and German, 1992). This locus has some similarity with the *Rsp* locus in that it is non-protein coding DNA and consists of tandem repeats of moderate length (*D1Z2* repeats are ≅40 bp each) arranged in clusters with intervening unique sequence DNA.

The results of Groden and German (1992) are summarized as follows. Clonal cell lines derived from a Bloom's syndrome patient lymphocyte show variability in the number of bands on a Southern blot. The exact number of tandem repeats is not quantified since the loss or gain of bands in this case may include unique sequence DNA segments. Both loss and gain of bands were observed. Bloom's syndrome cell lines also show abnormally high rates of intra- and interchromosomal exchange (Ray and German, 1983). These two observations were interpreted to suggest that the hypervariability is likely to be the result of unequal sister chromatid exchange due to mispairing of the tandem repeats. The mechanism of unequal sister

chromatid exchange (USCE) is described in detail in a subsequent section.

Durfy and Willard (1989 (cited in Harding et al 1991, interpretations are his)) examined a centromeric alpha-satellite tandem repeat DNA in the X chromosome of humans with a repeat unit size of 171 bp. Based on intra- and inter-sequence comparison, it was inferred that mutation occurred by USCE and that larger misalignments are uncommon, with shorter misalignments predominating.

Examples of longer repeat units producing length polymorphism in *Drosophila* genes include the following:

In the period (*per*) gene in *D. melanogaster*, there is length variation in the protein produced by a di-peptide repeat unit, a Thr-Gly pair (Costa et al, 1991). The alleles differ in length but are otherwise phenotypically normal. The variants most commonly seen differ by sets of 3 amino acid pairs, giving a repeat cassette of 18 nucleotides. The analogous locus in *D. pseudoobscura* also has length variation, evidently due to a 15 nucleotide repeat motif.

In the *dec-1* eggshell locus in *D. melanogaster*, length variation is due to an 78 nucleotide repeat unit (Andersson, 1993). The longest observed normal allele differed from the shortest by 12 copies (936 bp). In this case as well as in the *per* locus, there is no evidence of hypermutability but the similarity of inter-repeat sequence indicates that it is probably the product of sister chromatid exchange (USCE) or slippage amplification (RSR).

A similar comparison of inter-repeat DNA sequences in the *Rsp* locus has been done (Cabot, et al, 1993). The 240bp repeat unit actually a dimer consisting of two 120 bp units with slightly different sequences. Divergence between the left and right halves of the dimer can be identified by 16 diagnostic nucleotide differences. Examination of the sequences of monomers and trimers showed several cases of what appear to be hybrid sequences . This was interpreted as evidence of frequent unequal exchange. In the same study, comparison of sequence similarity of repeats within chromosomes versus between chromosomes shows that repeats on the same chromosome are no more similar than they are to those on different chromosomes. This may be explained either by frequent homologous exchange or by recent common ancestry of the locus. The later implies a recent selective (or meiotic) sweep for this chromosomal segment.

1.3.2 Theoretical Studies of HRTR

There are several recent theoretical studies that consider the dynamics of tandem repetitive DNA (Dover, 1995). Some recent publications of this type are summarized in the following.

Charlesworth, et al (1986) observed that highly repetitive DNA tends to be associated with regions of reduced recombination. (Human minisatellites are a contrary example (Jefferys et al, 1985; 1988)). This may simply be the result of such sequences being

removed by recombination in most regions but not in these. Theoretical analysis of the evolutionary equilibrium between genetic drift, natural selection and mutation pressure was based on assumptions that 1) mutation occurs by unequal exchange that is completely reciprocal, 2) that there is an inverse relationship between fitness and copy number. This analysis led to the further suggestion that stabilizing selection on the copy number of the tandem repeats may lead to suppression of recombination. Consequently, their suggestion is that the tandem repetitive DNA doesn't merely accumulate in the regions of reduced recombination but rather that the reduced recombination is a result of selection on the repetitive DNA.

Stephan (1986) derived explicit expressions for asymptotic behavior of tandem repetitive DNA under recombination in large populations versus small populations. The analysis was also supported by simulations. He confirmed the prediction that mean time to loss is highly dependent on recombination rate. He also came to the surprising conclusion that highly repetitive tandem repeats persist longer in smaller populations than in larger ones.

Another mechanism related to recombination that has been considered is intrastrand exchange. In this process, a loop forms in a DNA strand, bringing into alignment two repeats. An exchange produces a deletion. Walsh (1987) invoked intrastrand exchange as a source of bias in the change in number of repeats in a tandem array. This leads to a net tendency to decrease in number (most models of

unequal exchange assume no change in mean number of copies). The obvious conclusion is that persistence time is dependent on the ratio of intrastrand to interstrand exchange events. Specifically, a strand starting with (z +1) excess copies will have a mean persistence time = $z(1 + \gamma/\varepsilon)$ where γ is the interstrand rate and ε is the interstrand rate of exchange. Furthermore, if the assumption is made that selection imposes a lower bound (n) on copy number, then the upper bound of mean copy number is given by $n + 1/\alpha$ (where $\alpha = 2\varepsilon/\gamma$). Walsh concluded that expansion of arrays is unlikely unless $\alpha >> 1$, so some form of amplification (bias toward increase) must be operating in the evolution of large arrays. Replication slippage (RSR) was suggested as one possible mechanism for this. Another is a "rolling-circle" type of replication event in which intrastrand pairing during DNA replication leads to a duplication event. Sequence structure of human alphoid satellites suggests that this may occur (Deville et al, 1986; Jones and Potter, 1985) but its importance is not clear.

Stephan (1989) simulated a system of dynamic tandem repetitive DNA to explore what patterns of sequences are produced by different parameters. This simulation follows a single chromosomal lineage through time. The processes of USCE and RSR are modeled as biased toward amplification and sequence dependent, with probability of mispairing decreasing over time as random nucleotide substitutions occur. At each generation, after all mutational processes have occurred, one of the two daughter chromosomes is chosen at random to go on to the next generation. At

high recombination frequencies, the nucleotide patterns generated in the simulations are simple and highly regular, with short, nearly identical sequences repeated in tandem. Lower recombination rates lead to longer and more-complex repeat units with higher order periodicities.

Harding et al (1992) were interested in hypervariable human micro-and minisatellite tandem repeats primarily as tools for forensics, gene mapping and population genetics/evolutionary biology. To be useful in such a context, it is necessary to understand the evolutionary dynamics of the repeats (Rubinsztein et al, 1995; Tautz, 1989). In order to test specific assumptions, they compared previous analytical models (Gray and Jefferys, 1991; Walsh, 1987) with the results of the following simulations.

Harding et al (1992) used an iterative Monte Carlo simulation in which initial arrays of 2 repeats were subjected to random misalignments and recombination that produces shrinkage and expansion. When an array drops to one copy the simulation ceases, otherwise, a single allele is followed through many generations (up to millions). Comparisons of assumptions about recombination were made between simulations constraining misalignment in 3 different ways. First, misalignment is limited to a single repeat unit. Probability of exchange is 1 given a misalignment of 1, and 0 for any other misalignment. Second, misalignment is limited to a specific "maximum-target" length (t) and probability of exchange is uniform for all misalignments up to t, with larger misalignments prohibited.

In the third case, misalignment is allowed up to the maximum size of the allele (n) with a non-linear probability of exchange. The probability function for exchange is based on an empirical distribution for human minisatellite MS32 (Gray and Jefferys, 1991). This probability function makes small misalignments more likely than large ones, which was observed with the MS32 data. The probability of allelic recombination is a linear function of allele size.

They followed lineages to a specified number of repeat units (20, 50, 200, or 500). They defined 5 phases: lag, in which the lineage remains at 2 copies; gain, in which there is a rapid increase in copy number up to the target number; dwell, which is the period during which the array remains at or above the target; decay, when the array falls below the target and reduces to 2 repeats; end, when the array drops to one copy, i.e., extinction. They found the percentage of lineages reaching the targets was about the same for all three recombination rules. Furthermore, the percent reaching each target was proportional, with 4-5% reaching 20 and 0.4-0.5% reaching 200.

As in Walsh (1987), Harding et al (1992) concluded that persistence time for arrays depends critically on a balance between a moderate amount of amplification relative to a low rate of recombination. Further, the persistence time for arrays was dependent on the amount of constraint on misalignment. Tightly constrained systems can have slow rates of change while less constrained systems can be quite dynamic. In all cases, however, the dwell phase is far shorter than the expansion or contraction phases.

This underscores the observation that systems like these are dynamic and so alleles should not be expected to be stable.

Stephan and Cho (1994) extended Stephan's 1989 study. As in the earlier study, a single chromosome lineage is followed through time. Nucleotide substitutions occur at a low rate, according to a Poisson process. Both USCE and amplification by RSR occur, with mispairing being dependent on sequence similarity. Selection does not work on sequences *per se*, but rather on sequence length with any sequence longer than a specified threshold being eliminated.

It was found that, under the conditions of the model, repetitive sequences can be generated from sequences that are initially random. This was observed in the earlier study (Stephan, 1989) but not emphasized as here. This self-organization of complex systems is a general phenomenon for which there is a body of theory (Kauffman, 1993). Stephan and Cho (1994) analyzed their system with this approach and found it in concordance with their result that lower recombination rate and higher selection threshold, produce longer repeat units with higher inter-repeat variability. The effect of selection can be understood intuitively: sequences that are less similar will undergo USCE less often and so have a tendency to cross the higher selection threshold less often. They therefore maintain greater lengths than more similar sequences, which will undergo more USCE. At lower selection thresholds, it is the shorter (and concurrently more similar) sequences that will tend to persist.

Stephan and Cho (1994) examined existing data to see if their

predictions fit for microsatellite, minisatellite (both euchromatic - high recombination) and satellite DNA (heterochromatic - low recombination). Satellite and minisatellite sequences did fit the predicted correlation of recombination rate with sequence length and complexity of higher order structures. However, the data for microsatellite DNA did not fit. This may be because the short length of the microsatellite repeats inhibits mispairing for USCE, so that RSR alone is responsible for change in copy number.

The simulations reported in the current study are similar to these recent publications in that they consider the dynamics of the evolution of tandem repetitive DNA under various mutation-like processes. However, this work differs from the preceding in two significant ways.

First, while previous studies were interested in the molecular evolution of the sequences, we are interested in the population dynamics of existing alleles that can transmutate (point mutation is not considered). Thus we are following frequencies of alleles in a population rather than single chromosome lineages.

Secondly, in the previous studies, selection takes the form of an upper or lower bound on copy repeat number. The action of *Sd* in our system represents a frequency dependent selective force, which adds a higher level of complexity to the model. In fact, we are specifically interested in the system from the point of view of the effect of the Responder locus on the frequency of *Sd* alleles.

Thus, the recent theoretical studies on the evolution of tandem repetitive DNA provide additional refinement to the simulations but do not address the fundamental questions we are asking.

1.3.3 Possible Mechanisms of Generation of Highly Repetitive Tandem Repeat Structure

1.3.3.1 Unequal Sister Chromatin Exchange

During prophase I of meiosis, chiasmata form and reciprocal exchange occurs between homologous chromosomes. Chiasmata also form between sister chromatids, and this is believed to play a role in maintaining cohesion of the tetrads until the metaphase I / anaphase I transition (Miyazaki and Orr-Weaver, 1992). Ordinarily, any exchange of genetic material between sister chromatids has no effect, since the chromatids are identical. However, when the sister chromatids contain tandemly repeated DNA sequences, the possibility for misalignment exists. This provides a potential mechanism for expansion or contraction of repeat copy number.

In the Unequal Sister Chromatin Exchange (USCE) model, sister chromatids synapse out of register at simple repeat structures and recombine. One strand then has extra copies while the other strand has exactly the same number of copies less. This model has been explored extensively (Kruger, 1975; Ohta, 1980; Perelson, 1977; Smith, 1974; 1976; Stephan, 1986; 1987; Stephan and Cho, 1994;

Walsh, 1987). Without selective constraints on upper and lower bounds on the number of copies, the model predicts fixation at one copy of the repeat unit.

The USCE model is appropriate for the present study, based on the observation that since *Rsp* is made up of a variable number of repeats of 240 bp each, it may be subject to events of unequal sister chromatid exchange (USCE) that would lead to reduced or increased number of repeats on a given chromosome, as suggested by preliminary work done by Wu and Takahata (unpublished). A misalignment during normal chromosome recombination results in symmetrical increase and decrease respectively in copy number in the two sister chromatids involved. This is illustrated in Figure 1.

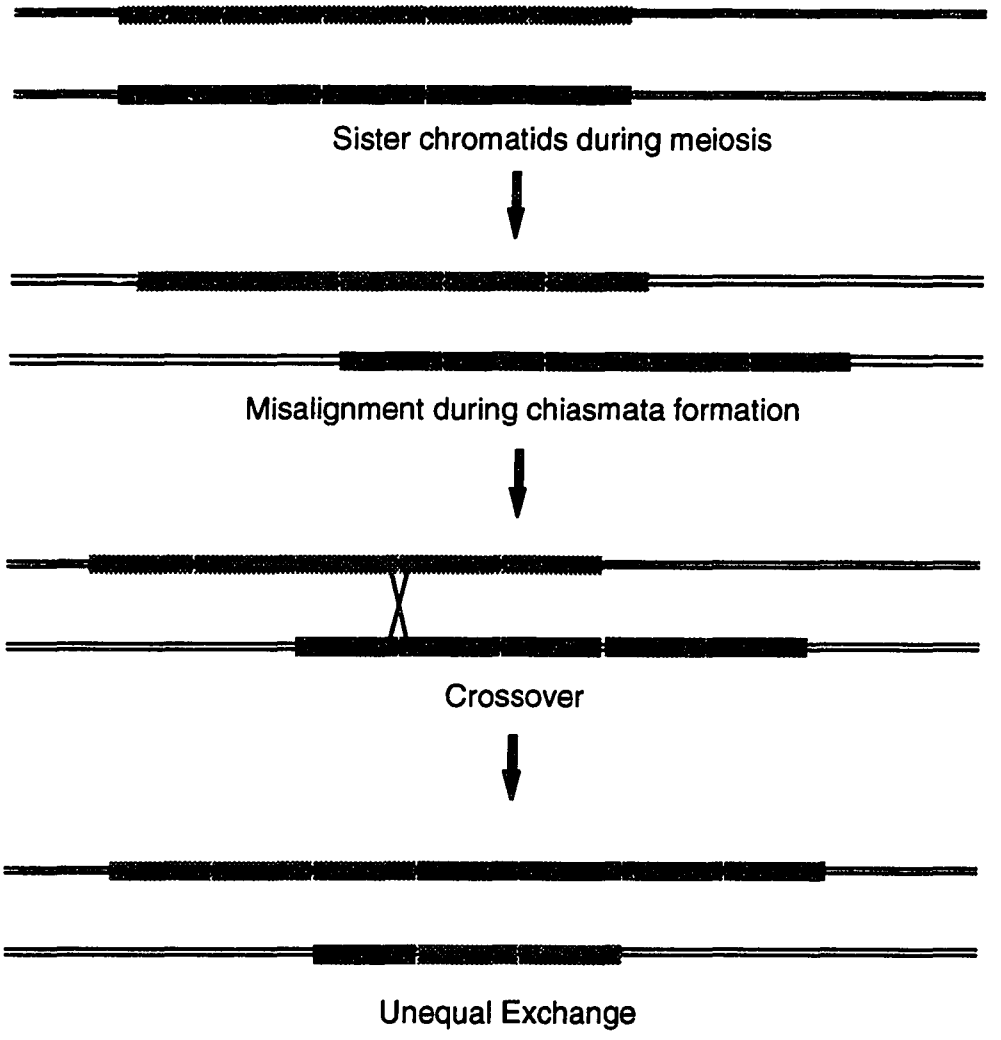


Figure 1. Schematic representation of the process of Unequal Sister Chromatid Exchange.

Unequal sister chromatid exchange is known to produce duplications and deletions (Peterson and Laughnan, 1963 cited in Ritossa, 1976; Roberts, 1976). Although interchromosomal exchange does not normally occur in males, intrachromosomal exchange does occur in males as well as in females, albeit at a frequency roughly 10-fold less than ordinary meiotic recombination (Ashburner, 1989 Chapter 17 and citations therein; Roberts, 1976). There is also evidence (Gatti et al, 1979) that USCE is more common at euchromatin-heterochromatin boundaries and in heterochromatin, which is the location of Responder. The pattern of sequence similarity among repeat units within and between Responder chromosomes has been interpreted as evidence of a high rate of USCE (Cabot et. al., 1993; Wu and Hammer, 1990). In at least one instance, a Rsp^s that had been made Rsp^i by deletion of most of its repeat units reverted to Rsp^s (Ganetzky, 1977), apparently by amplification of the residual repeat arrays (Wu and Hammer, 1990). The observations make USCE a likely candidate for a mechanism to increase the number of repeat units in the face of selection against high numbers by drive.

1.3.3.2 Replication Slippage Repair

Another proposed mechanism is the replication-slippage-repair (Levinson and Gutman, 1987; McMurray, 1995; Schlotter and Tautz,

1992; Strand, et al, 1993; Tachida and Itzuka, 1992). In this model (also known as mismatch repair), DNA polymerase slips during replication of a tract of DNA so that the template and primer strands temporarily disassociate. If the region has simple repeat structures, then reassociation in a misaligned configuration is not unlikely. Formation of hairpin loops may also contribute to misalignment. Various mechanistic models of this type of event are treated in detail in McMurray (1995), but since they all produce approximately the same result, a simplified version is sufficient here. The process is illustrated in Figure 2.

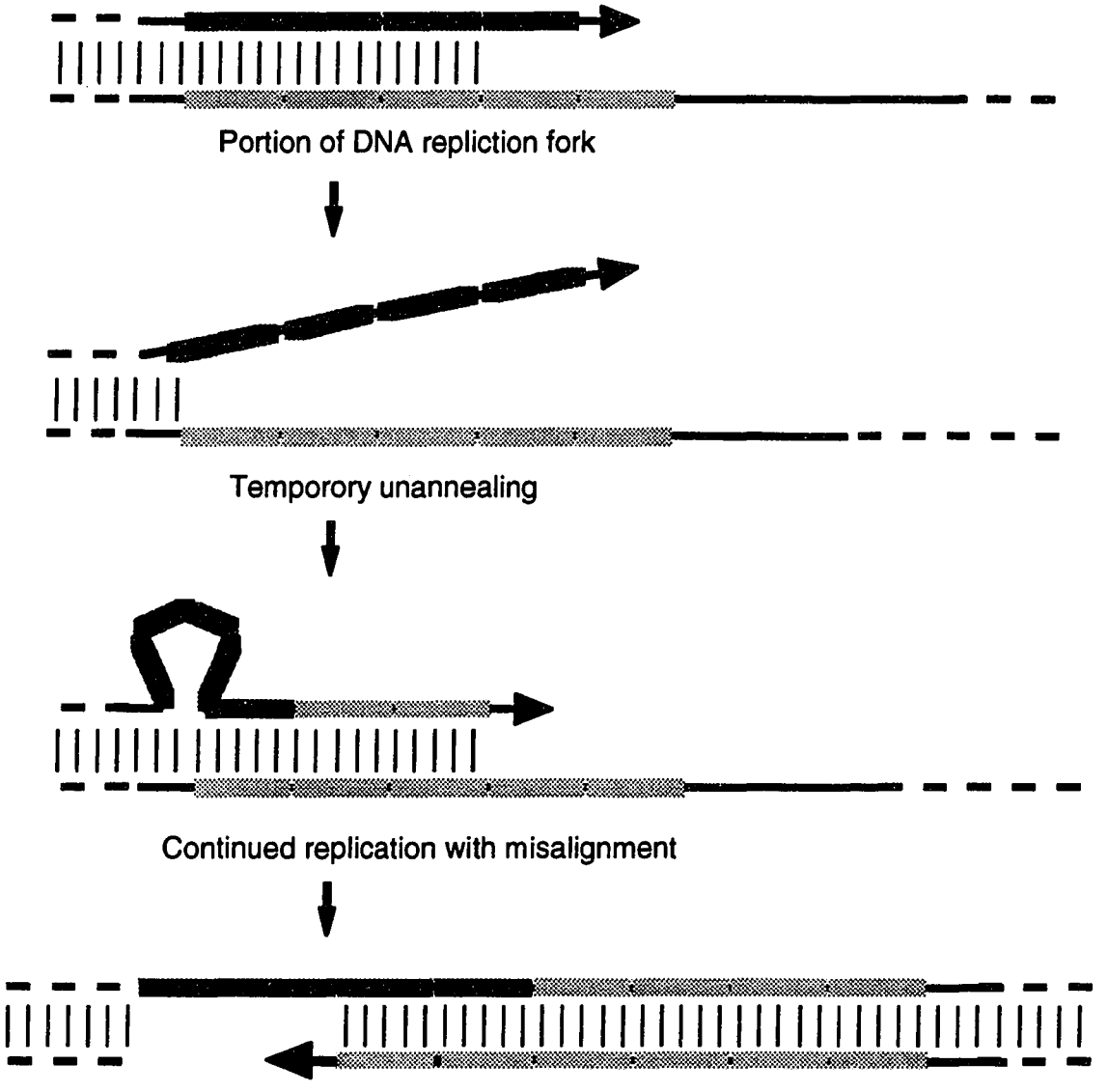


Figure 2. Schematic representation of the process of Replication Slippage Repair

If the unpaired bases are in the template strand, this results in a deletion in one of the strands. If the unpaired bases are in the primer strand, continued synthesis results in a duplication in one strand with no corresponding decrease in the other strand. In either case, DNA mismatch repair mechanisms then fill in the unpaired bases. Studies in bacteria and yeast (Strand, et al, 1993) suggest that either the 2 possible loops (template versus primer strand) do not occur with equal frequency or that the repair mechanisms are not equally efficient on the 2 strands with the net result being that there is a tendency to increase number of copies.

Strand et al (1993) examined the stability the length of a tract of dinucleotide repeats in yeast. They interpreted the effects of several DNA replication and proofreading mutants as indication that replication slippage is responsible for observed changes in the size of the repeat tract. They measured rates of tract alteration in wild type strains of 1.3×10^{-5} per mitotic cell division and 3.2×10^{-5} per meiotic cell division. The rates were increased 100 to 1000 fold by mismatch repair mutants.

Based on observed length variation in 5 human microsatellite sequences, Jefferys et al (1988) estimated a mutation rate of up to 5% per gamete. In the most unstable locus, the length variations are due to gain or loss of variable numbers (4-200) of a 9 bp repeat unit. Based on all 5 loci, they also found a weak statistical relationship between length of repeat and mutation rate. This direct

proportionality is a prediction that all the out-of-register alignment models share.

Human sperm go through approximately 400 cell divisions while the oocytes undergo approximately 24 divisions. A replication slippage model predicts that rate should be proportional to number of mitotic divisions while a sister chromatids exchange model predicts that only number of meioses determines rate. Jefferys et al (1991) found that the mutation rate was indistinguishable in sperm and oocytes. This would appear to support an exchange-mediated process. However, sperm had an excess of changes of small size compared to oocytes. A high rate of small changes is predicted if they are generated by replication slippage events. It may be that one process predominates in sperm and the other in oocytes.

Because of the usefulness of minisatellite DNA's in studies of human evolution and forensics, there is considerable interest in their mutational dynamics (Gray and Jefferys, 1991; Valdes et al , 1993). The general result of these studies is that for humans, the observed frequency distributions could be generated by a stepwise mutational model, with parameters reasonable for humans, that does not require asymmetrical amplification. The variance of the distribution was found to be a good estimator of $N\mu$, where μ is the mutation rate and N the effective population size. While the generality that single copy is an absorbing state and so all alleles should go to fixation at 1 is still true, the number of generations since the emergence of humans

is sufficiently short that the observed frequencies are predicted as possible transient states.

Harding et al (1991) are of the opinion that for shorter sequences, replication slippage is relatively common but that such slippage is less likely to occur with longer sequences. Therefore, for tandem repeats with units on the order of 100 bp, USCE may be the major mechanism at work while RSR dominates for shorter sequences (such as trinucleotides). This is supported by the results of Stephan and Cho (1994), summarized above.

1.3.3.3 Double Stranded Break Repair

Another likely biological mechanism for expansion of a set of tandem repeats is found in the process of repair of double strand break repair (Thompson-Stewart et al, 1994; Engels, et al, 1990). A number of molecular events can produce breaks in double stranded DNA. A prominent example is P-element excision (Gloor et al, 1991; Engels, 1989), in which the P-element excises itself with high precision, leaving a double stranded break. Other inverted-repeat transposons, including *Ac*, *hobo* and *mariner* are believed to transpose by similar mechanisms (Berg and Howe, 1989).

In this model, such a break is expanded into a double stranded gap by exonuclease action. This lesion is then repaired using homologous as a template. The repair template is invaded by a single stranded end based on a very short stretch of sequence homology

(Gloor, 1991). With repetitive sequence, alignment can be made out of register (Figure 3). When the repair continues, extra repeat copies will be produced. This mechanism would always result in an increase in copy number rather than any decrease. Therefore, this model predicts a bias toward increase in copy number.

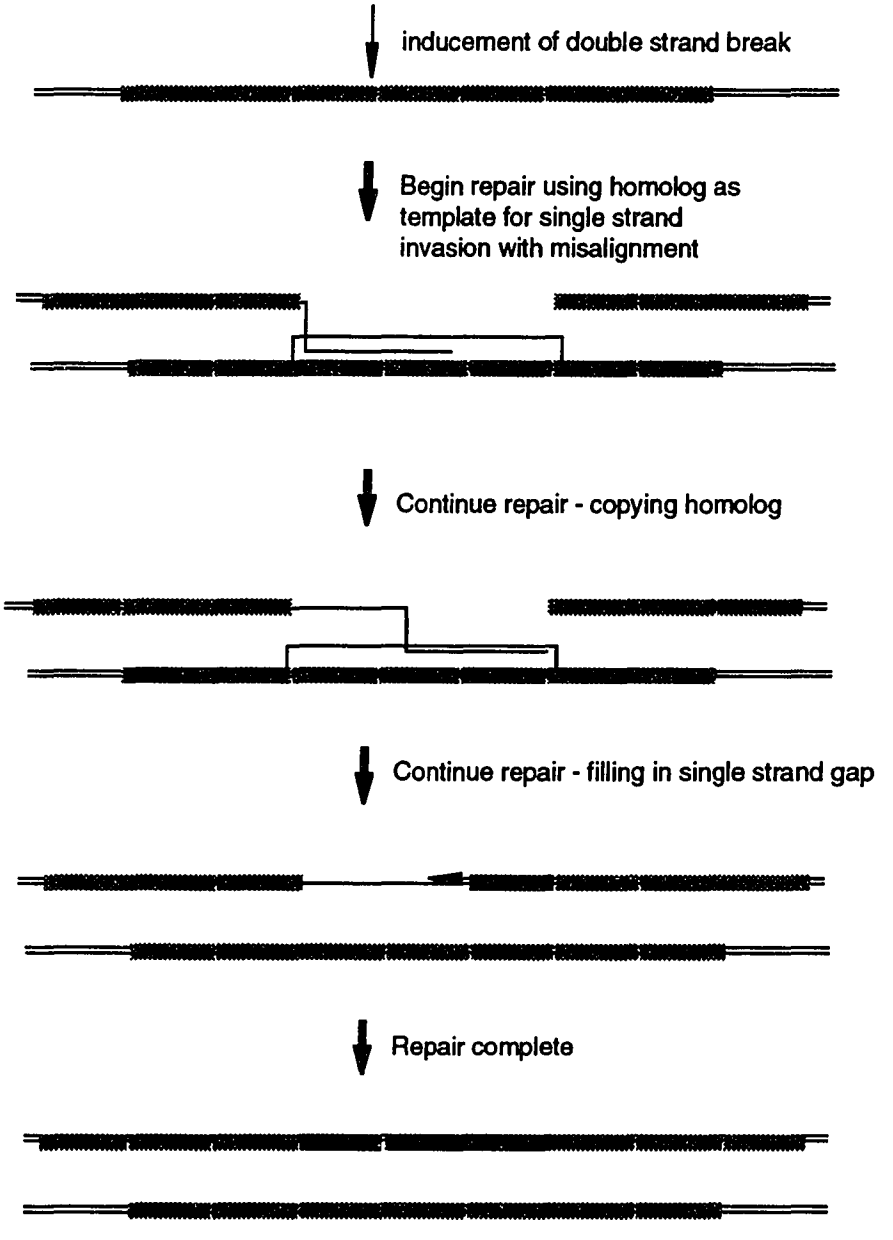


Figure 3. Schematic representation of the process of Double Strand Break Repair

.

The DSBR model differs from the USCE and RSR models above in that it produces only increases in copy number. Furthermore, as modeled here, the probability of expansion is not length dependent and so the rate of change is constant. The relative frequency of sensitives and insensitives is ultimately determined by a balance between the rate of conversion of insensitives into sensitives and negative selection, in the present case, by *Sd* as determined by the strength of the meiotic drive.

1.3.3.4 Mutation Rate as a function of SD frequency

This model is based on a hypothetical interaction of the drive locus and the insensitive Responder so that the level of sensitivity increases, due to duplication of the repeats by the action of *Sd*. Conceptually, *Sd* produces some product that affects the Responder locus during each meiotic round by causing an increase in the number of repeats. A chromosome has a likelyhood of being rendered dysfunctional in gametogenesis that is proportionate to the number of repeats it carries. This is manifested as meiotic drive, as a proportion of the target chromsomes in an individual, determined by the number of repeats, fail to undergo gametogenesis properly. Chromsomes that do not become dysfunctional undergo normal gametogenesis but increase in copy number and so become more likely to reach a critical number of repeats in the next generation. That is, they become more sensitive to meiotic drive.

The hypothesized process is unidirectional, with only expansions occurring. Sensitivity grows over time, but at a rate dependent on the frequency of *Sd*. The relative frequency of sensitives and insensitives and the frequency of *Sd* is determined by a balance between the rate of conversion of insensitives into sensitives and the strength of drive.

There is some evidence in support of this model. The SD system has some poorly understood associated phenomena (Hartl and Hirazumi, 1976) and the results of some experiments have not been explained in light of the current theories of SD action.

In an early study, Crow et al (1962) hypothesized that if SD induces a physical break in its homolog as part of the action of distortion, then males in which drive is active should show a higher recombination rate than males without drive when irradiated, which should provide the second break needed for recombination. There was, in fact, a two-fold increase in the amount of recombination in the appropriate genotypes, as predicted. This was interpreted at the time to mean that SD did induce a break in its homolog. The breakage hypothesis for the action of SD has since been rejected on other grounds but the results of this experiment have not been satisfactorily explained (Hartl and Hiraizumi, 1976).

Another early experiment (Hiraizumi, 1961) suggests that SD may induce mutations in its homolog. Homozygous viability of *cnbw* chromosomes derived from SD/*cnbw* males was tested and it was found that there was a decrease in the viability of these

homozygotes. Three of the 148 chromosomes examined had recessive lethals versus none among the 86 control chromosomes. Furthermore, the lethals all mapped to within 5% of the SD region.

These results suggest that SD is doing something to change the DNA of its homolog, although it is not at all clear what it is. While there is no direct evidence that SD causes an increase in the number of repeats at the *Rsp* locus, the model is plausible. It is interesting for the present study, because it involves a mutational process that is under the control of the driving locus.

Chapter 2
Materials and Methods

2.1 Population Simulations

2.1.1 General Comments on Simulation Methods and Assumptions

Segregation Distorter (SD) is a system of two loci with two alleles each. However, the level of recombination between *Sd* and Responder is extremely low and so they are often referred to as haplotypes. For the purpose of these simulations, recombination can be ignored and the haplotypes treated as alleles.

The basic haplotypes are *Sd-Rsp^i* (*SD*), *Sd-Rsp^s* (Suicide), *Sd$^+$-Rsp^i* (insensitive) and *Sd$^+$-Rsp^s* (sensitive). Since the *Sd-Rsp^s* (Suicide) combination only arises by recombination and is extremely maladapted, it is excluded from this study. The *Sd-Rsp^i* combination is designated *SD* and treated as immutable.

Among *SD$^+$* chromosomes, the number of repeats is known to range from a few (tens) to several thousand. Directly modeling this would be intractable (the memory and processing time required to follow several thousand allele classes would be beyond most mainframe computers) and is not really necessary to follow the dynamics of the system. In order to more realistically simulate the

process of gradual change in *Rsp* repeat copy number, it was necessary to model the *Rsp* chromosomes as containing blocks of repeats. Each of these blocks, designated R, corresponds to a few tens of repeats. In fact, the real difference between functional classes (alleles with different sensitivities) is very roughly logarithmic (10^1 = ins, 10^2 = semisens, 10^3 = sens, 10^4 = supersens, etc.). Therefore, a single block makes the R chromosome like the true Rsp^i found in nature, which contain between 10 to 30 copies of the repeat structure, with no sensitivity to distortion. Increasing numbers of repeat blocks correspond to the range of copy number and sensitivities found in nature. A chromosome bearing two blocks, RR, is equivalent to a semisensitive chromosome with 50 to 100 repeat copies. A chromosome bearing three blocks, RRR, is equivalent to a classical sensitive chromosome with a few hundred repeat copies. A chromosome bearing four blocks, RRRR, is equivalent to a very sensitive chromosome with several hundred to thousands of repeat copies. A chromosome bearing five blocks, RRRRR, is equivalent to a supersensitive chromosome with thousands of repeat copies. This approach has an additional advantage since four functional classes of Responder were defined (Martin and Hiraizumi, 1979; Temin and Marthas, 1984) and data on sensitivity of these four broad classes is available. Such information is necessary to parameterize meiotic drive for the simulation. Thus, the 6 haplotypes used in this study are *Sd*, R, RR, RRR, RRRR and RRRRR.

The computer simulation goes through a multistep process of mutational conversion followed by meiosis with meiotic drive, iterated for a specified number of generations. The actual procedure used for conversion depends on the mutational model specified.

The simulation has three further refinements: (1) since drive occurs only in males, new allele frequencies are calculated separately in males and females and then combined, (2) a factor was added to allow for inclusion of reduced fertility of SD/SD homozygous males, and (3) a procedure using a Probit transformation to calculate the strength of drive against various alleles. This is necessary since the effect of multiple doses of Responder is not linear with respect to *k*, but is linear with respect to the probability of reaching a threshold of liability to sperm dysfunction (Lyttle, 1986). This probability is derived from k by the Probit transformation (Beyer, 1968).

The basic drive specified, k, is used as the drive against RRR (standard sensitive). An empirical curve based on measurements of sensitivity or chromosomes from nature (Martin and Hiraizumi, 1979; Temin and Marthas, 1984) was used to estimate drive strengths of *Sd* against the semisensitive (RR), very sensitive (RRRR) and supersensitive (RRRRR) chromosomes.

The user inputs values for strength of drive (k), insensitive to sensitive conversion rate (PU), number of generations to run (gens), fertility of SD/SD homozygote males (fert), and initial allele frequencies (x1-6). Choosing the initial frequencies makes it possible to simulate the system under various conditions, such as the origin of

the system when one (or more) new mutants first arose and so were in very low frequency.

In general, the procedure "conversion" changes the number of Responder blocks at a specified rate that is analogous to an ordinary mutation rate. The actual rate is dependent on the specific model. This rate is generally several orders of magnitude higher than an ordinary point mutation rate because it is intended to simulate a directed, rather than random, process. The conversion rate is a probability of conversion from one allele to another in one generation.

The procedure "meiosis" takes the allele frequencies after conversion, combines them in Hardy-Weinberg proportions, modifies them by the effects of drive, and of fertility reduction and then outputs the frequencies at time t + 1.

The program is designed so that allele frequencies over time are stored in a matrix with dimensions of number of alleles (6) by the number of generations of the simulation for ease of manipulation and for storage as a file. The upper limit on number of generations is about 400 due to computer memory constraints. In practice, this limit is not a problem since for most runs, the frequencies stabilize within 20 to 50 generations.

A second version of the program does not use a matrix and so can be run for any number of generations, but it has the disadvantage of not being able to store or recall values of previous generations. This version of the program automatically steps through

incremental values for k and PU and saves the results to files so that a large number of combinations can be easily tested.

For purposes of graphical presentation of the results, three values of k were chosen, representing high drive (k=1.0), moderate drive (k=0.85) and low drive (k=0.7). Ten values of PU were chosen to cover the range of mutation rates that were of interest in this study. These were: 1×10^{-1}, 5×10^{-2}, 1×10^{-2}, 5×10^{-3}, 1×10^{-3}, 5×10^{-4}, 1×10^{-4}, 1×10^{-5}, 1×10^{-6}, 1×10^{-7}. Values lower than 1×10^{-7} gave rates of change too small to be of interest and values greater than 1×10^{-1} were not considered to be biologically meaningful.

2.1.2 Models of Mutational Mechanisms

2.1.2.1 Double Strand Break Repair Model

The procedure for the double strand break repair model simulates the change in allele frequencies due to the mutation procedure using the following set of equations. Xi_t is the frequency of the i^{th} allele at time t and PU is the probability of a conversion event occurring. X1 through X6 are the alleles R, RR, RRR, RRRR, RRRRR and *Sd*, in order.

$$X1_t = (1 - PU)\, X1_{t-1}$$

$$X2_t = (1 - PU)\, X2_{t-1} + PU\, X1_{t-1}$$

$$X3_t = (1 - PU)\ X3_{t-1} + PU\ X2_{t-1}$$

$$X4_t = (1 - PU)\ X4_{t-1} + PU\ X3_{t-1}$$

$$X5_t = (1 - PU)\ X5_{t-1} + PU\ X4_{t-1}$$

$$X6_t = X6_{t-1}$$

The resulting values then enter the meiosis procedure, where meiotic drive and adjustment for fertility reduction of the SD/SD homozygotes occurs. In the following set of equations, F is the fertility of the SD/SD homozygote, k is the basic drive strength against the sensitive (X3 = RRR), hk is the drive strength against the semisensitive (X2 = RR), kk is the drive strength against the very sensitive (X4 = RRRR), and kkk is the drive strength against the supersensitive (X5 = RRRRR). These values are calculated from an empirical curve based on measured sensitivities of various Responder alleles in the literature (Martin and Hiraizumi, 1979; Temin and Marthas, 1984).

$$X1_t = X1_t^2 + X1_t\ X2_t + X1_t\ X3_t + X1_t\ X4_t + X1_t\ X5_t + X1_t\ X6_t$$

$$X2_t = X2_t^2 + X2_t\ X1_t + X2_t\ X3_t + X2_t\ X4_t + X2_t\ X5_t + 2\ X2_t\ X6_t\ (1 - hk)$$

$$X3_t = X3_t^2 + X3_t\ X1_t + X3_t\ X1_t + X3_t\ X4_t + X3_t\ X5_t + 2\ X3_t\ X6_t\ (1 - k)$$

$$X4_t = X4_t^2 + X4_t\ X1_t + X4_t\ X2_t + X4_t\ X3_t + X4_t\ X5_t + 2\ X4_t\ X6_t\ (1 - kk)$$

$$X5_t = X5_t^2 + X5_t\ X1_t + X5_t\ X2_t + X5_t\ X3_t + X5_t\ X4_t + 2\ X5_t\ X6_t\ (1 - kkk)$$

$$X6_t = F\ X6_t^2 + X6_t\ X1_t + 2\ hk\ X6_t\ X2_t + 2\ kk\ X6_t\ X3_t + 2\ kk\ X6_t\ X4_t + 2\ kkk\ X6_t\ X5_t$$

These frequencies are adjusted for losses due to fertility reduction in SD/SD homozygotes by dividing by the sum of frequencies.

The procedure is iterated through several generations and the results stored in a two dimensional matrix. The results are plotted on a line graph of time in generations versus frequencies. The procedures for meiosis, data storage and graphing are the same in all the simulations.

2.1.2.2 Mutation as a Function of SD Frequency

The procedure for the *Sd* mediated mutation model (Hits) simulates increases in sensitivity as being induced when an *Rsp* chromosome finds itself with an *SD* chromosome. It is therefore a function of their frequencies. Change is dependent on a "hit rate" that simulates the number of times the insensitive must encounter an *Sd* to be converted to the next higher level of sensitivity (i.e., reach a threshold in number of repeats added by the action of *Sd*). For ease of programming, instead of tracking individual chromosomes and keeping track of number of times each occurs with an *Sd* (not possible, given an assumption of infinite population size), hit rate is treated as an average probability of being converted and is equal to the inverse of the number of hits to convert. The mutation rate (PU) is therefore a probability of change in a single event. Variables are as described above.

$$X1_t = X1_{t-1} - PU\ X6_{t-1}\ X1_{t-1}$$

$$X2_t = X2_{t-1} + PU\ X6_{t-1}\ X1_{t-1} - PU\ X6_{t-1}\ X2_{t-1}$$

$$X3_t = X3_{t-1} + PU\ X6_{t-1}\ X2_{t-1} - PU\ X6_{t-1}\ X3_{t-1}$$

$$X4_t = X4_{t-1} + PU\ X6_{t-1}\ X3_{t-1} - PU\ X6_{t-1}\ X4_{t-1}$$

$$X5_t = X5_{t-1} + PU\ X6_{t-1}\ X4_{t-1}$$

$$X6_t = X6_{t-1}$$

These values then undergo the meiosis process. The procedures for meiosis, data storage and graphing are as described previously.

2.1.2.3 Unequal Sister Chromatid Exchange

In the basic USCE algorithm, PU represents the probability of an unequal sister chromatid exchange event occurring. When the event occurs, there is a reciprocal exchange of repeats so that one of the chromatids gains the same number of repeats that the other loses. The number of repeats exchanged is dependent on the size of the repeat. PU specifies the frequency with which an event occurs. When it occurs, the probability of each degree of mismatch is evenly distributed among all possibilities. For example, an unequal exchange event between the sister chromatids of an RRRR would produce misalignments of one, two or three repeat blocks but one and three produce the same result, so that outcome is twice as likely as the outcome of a misalignment of two repeat blocks, which only occurs one way. The two versions of the simulation described below differ

in how USCE events that would result in alleles with less than one repeat block or more than five are resolved.

2.1.2.3.1 USCE Truncation

This algorithm is designed to simulate USCE in which alleles that grow too small or too large are lost. When a USCE occurs involving an R allele, 25% change to RR, 50% remain unchanged and 25% is lost. When USCE occurs to an RRRR or an RRRRR, that proportion that would become RRRRRR or larger is lost. The resulting allele frequencies are adjusted to a total frequency of 1.0 to account for that proportion lost.

There are two versions of this model that differ in how they handle the ends of the distribution: Truncation and Bounceback.

In the first version of the simulation, a single copy of the repeat is an absorbing state since it can not mispair to produce higher numbers. If this is so, then the result will always be fixation at 1 copy. Some simulation studies (e.g., Stephan, 1989) have handled this by assuming a selection against less than two copies (i.e., truncating the distribution). These simulations make the assumption that each repeat unit (R) is not really just one copy of the repeat but rather a small number, i.e., the difference between R and RR is not just 240bp but is really several tens of copies at least. Given this, it is possible for an R allele to misalign and become RR and "0" (no R's). In one version of the simulation, R's undergoing USCE may stay R

(50%), or undergo USCE to become RR (25%) and 0 (25%). The latter class is lost. This is referred to as the USCE truncation model. At the other end of the distribution, RRRRR produces a similar problem. Many simulation studies create a truncation selection on alleles with too many copies. The USCE truncation model truncates any allele that grows to more than 5 copies in the same way as 0's above.

$$X1_t = (1 - PU)\, X1_{t-1} + PU\,(0.50\, X1_{t-1} + 0.25\, X2_{t-1} + 0.11111\, X3_{t-1} + 0.0625\, X4_{t-1} + 0.04\, X5_{t-1})$$

$$X2_t = (1 - PU)\, X2_{t-1} + PU\,(0.25\, X1_{t-1} + 0.5\, X2_{t-1} + 0.22222\, X3_{t-1} + 0.125\, X4_{t-1} + 0.08\, X5_{t-1})$$

$$X3_t = (1 - PU)\, X3_{t-1} + PU\,(0.25\, X2_{t-1} + 0.33333\, X3_{t-1} + 0.1875\, X4_{t-1} + 0.12\, X5_{t-1})$$

$$X4_t = (1 - PU)\, X4_{t-1} + PU\,(0.22222\, X3_{t-1} + 0.25\, X4_{t-1} + 0.16\, X5_{t-1})$$

$$X5_t = (1 - PU)\, X5_{t-1} + PU\,(0.11111\, X3_{t-1} + 0.1875\, X4_{t-1} + 0.2\, X5_{t-1})$$

$$X6_t = X6_{t-1}$$

The procedures for meiosis, data storage and graphing are as described previously

2.1.2.3.2 USCE Bounceback

In this version of the simulation, R alleles undergoing USCE may stay R (75%) or become RR (25%) but never enter the "0" class. This is designed to simulate the likely case that most misalignments

would not be precisely at the end of a set of repeat units so that there would usually be a few copies left after an exchange between R's. These copies would then be available for misalignment and exchange in a later round of meiosis. This will be referred to as the USCE bounceback model. The approach for large alleles is reached by the following logic. After an allele gets a high copy number (e.g., >10), most of the exchanges that take place will result in new alleles that have >5 copies and so are still functionally sensitive. To implement this, the simulation tracks the proportion of RRRRR's that undergo USCE but don't become a smaller number and treats them as less likely to undergo a change to another allelic state in subsequent generations. That is, for USCE purposes, they have many more copies but for drive calculation, they are RRRRR's. This has the net result of making RRRRR a partial absorbing state, but with a "bounce-back" property similar to what happens to R's.

The equations are the same as in the truncation model, with the additions described above. The procedures for meiosis, data storage and graphing are as described previously.

2.1.2.4 Replication Slippage Repair

The RSR model simulates a process in which the repeated DNA sequences expand or contract by a process of slippage during replication. If the nascent strand and the template strand separate temporarily during replication misalignment, may occur among the multiple copies of the sequence. There is evidence that the process

may be asymmetrical, with expansions much more common that contractions. Since expansions are the process of interest in this study, the simulation considers the result of slippage to be entirely biased in favor of expansions.

There are 2 versions of this model that differ in the length of misalignments allowed: RSRsrm and RSRamm.

2.1.1.4.1 RSRsrm

In the first, called RSR-single repeat misalignment (RSRsrm), the number of repeats by which the strands may misalign is limited to one. Thus, three copies may become four in a single event but never five. As discussed above, the process allows only expansions. Large and small alleles are handled as in the USCE bounceback model, i.e., the are absorbing states with bounceback properties.

$$X1_t = (1 - PU)\, X1_{t-1}$$

$$X2_t = (1 - PU)\, X2_{t-1} + PU\ X1_{t-1}$$

$$X3_t = (1 - PU)\, X3_{t-1} + PU\, X2_{t-1}$$

$$X4_t = (1 - PU)\, X4_{t-1} + PU\, X3_{t-1}$$

$$X5_t = (1 - PU)\, X5_{t-1} + PU\, X4_{t-1}$$

$$X6_t = X6_{t-1}$$

The procedures for meiosis, data storage and graphing are as described previously.

2.1.1.4.2 RSRamm

In the second model, called allele-maximum misalignment (RSRamm), the degree of misalignment is determined by a probability that is a function of the length of the sequence, with small alignments being more likely than large ones. The probability function is an empirical one produced by Gray and Jefferys (1991) based on recombinational dynamics of the minisatellite MS32. The function is

$$Pr(i) = \frac{(n-i)^{3.4}}{\sum_{j=1}^{j=n-1}((n-i)^{3.4})}$$

where j is the size of the misalignment and i the point of exchange.

The equations are the same as in RSRsrm except for the probabilities of change for the larger alleles. Here, the process allows only expansions.

$$X1_t = (1 - PU)\ X1_{t-1}$$

$$X2_t = (1 - PU)\ X2_{t-1} + PU\ X1_{t-1}$$

$$X3_t = (1 - PU)\ X3_{t-1} + PU\ X2_{t-1}$$

$$X4_t = (1 - PU)\ X4_{t-1} + PU\ 0.91347\ X3_{t-1}$$

$$X5_t = (1 - PU)\ X5_{t-1} + PU\ (0.08653\ X3_{t-1} + 0.78382\ X4_{t-1})$$

$$overX5_t = PU\ (X5_{t-1} + 0.21618\ X4_{t-1}) + overX5_{t-1}$$

$$X6_t = X6_{t-1}$$

The procedures for meiosis, data storage and graphing are as described previously.

2.2 Molecular Techniques

The following are a brief descriptions of the techniques used. Detailed molecular protocols are in the Appendix.

2.2.1 Plasmid Constructs: PGN156 and H_0

The PGN156 plasmid used in this study was provided by Pat Powers of University of Wisconsin. Plasmid PGN156 contains a 5 Kb EcoR1 fragment in a pGeml vector (from Promega-Biotech). The 5kb EcoRl fragment includes the Sd^+ region and is unique to *Sd* and hybridizes to both the *Sd* and Sd^+ alleles from genomic DNA.

The H_0 plasmid used in this study was provided by C-I Wu. Plasmid H_0 is a 2.5 Kb EcoR1 fragment in a pUC9 vector (Wu, et al 1988).

2.2.2 Genomic DNA Extraction

2.2.2.1 Genomic DNA Extraction from Multiple Flies

Genomic DNA from multiple flies for characterization of isolines was extracted by Lifton protocol, as modified by Loretta Arcangeli for Dr. Haymer's lab.

2.2.2.2 Genomic DNA Extraction from Individual Flies

Genomic DNA from individual flies was extracted via PureGene Animal Tissue (Gentra Systems, Inc.) protocol modified for single flies.

2.2.3 Digoxigenin Labeled Probes

Plasmids PGN156 and H_0 were used to make randomly labeled digoxigenin-dUTP probes using the Genius DNA labeling Kit (Genius) with slight modifications.

2.2.4 Procedure for Squash Blots

Squash blots followed the protocol of Martin (1990). Whole flies were squashed on Magnagraph nylon. The membrane was then denatured for one hour, neutralized and fixed to the membrane by baking.

2.2.5 Southern Blot

Agarose gel electrophoresis and Southern blotting was done using standard protocols (Sambrook et. al., 1993) with some modifications. Samples mixed with loading buffer were loaded

submarine in TBE buffered gels with 0.7% agarose. These were electrophoresised for several hours, at 40-70 volts, until the loading dye reached the end of the gel. The gel was denatured for one hour, neutralized and then transferred to Magnagraph nylon, using 4x SSC buffer and blotting overnight.

2.2.5.1 Hybridization of the Membrane

The DNA-bound filter was prehybridized with 50% formamide hybridization solution in a sealable bag. Hybridization mix was 7 ul of labeled DNA (approximately 5-25 ng/ul) in TE added to 7 ml of the hybridization solution. The filter was incubated at 42° C overnight.

2.2.5.2 Post Hybridization Washes and Visualization

The post hybridization washes were high stringency (68° C). Detection was by the Genius Lumiphos Nonradioactive DNA Labeling and Detection Kit protocol. Visualization was by exposure to X-ray film for as little as 5 minutes or up to several hours as needed.

Sd genotype and Responder genotype data were collected from these blots.

2.3 *Sd* and *Rsp* Allele Frequencies via Southern Blotting

Total genomic DNA was extracted from single flies were digested with EcoRI and probed with dig-labeled PGN156 insert to detect *Sd* allele frequencies. Total genomic DNA was extracted from single flies were digested with MspI and probed with dig-labeled H_0 insert to detect Responder allele frequencies.

2.3.1 *Sd* Allele Frequency via Southern Blot

Sd frequencies were assessed using Southern blot of restriction endonuclease digested genomic DNA. The probe clone, PGN156, contains a 5 Kb EcoR1 that includes the Sd^+ region, is unique to *Sd*, and hybridizes to both the *Sd* and Sd^+ alleles from genomic DNA. The *Sd* and Sd^+ alleles differ in that the *Sd* allele contains an extra 5 Kb and releases an EcoRI fragment of 12 Kb while the Sd^+ allele releases an EcoRI fragment of 7 Kb.

This size difference makes it possible to identify which alleles are carried by each individual. Genomic DNA was isolated, digested with EcoR1, separated by size on an 0.7% agarose gel, Southern blotted and visualized by hybridizing the blot with dig-labeled PGN156 probe DNA. Homozygotes show a single band of either 7 Kb or a 12 Kb, while heterozygotes show both bands. Thus, the homozygote and heterozygote frequencies were estimated by direct count of alleles in population samples.

2.3.2 *Rsp* Allele Frequency via Southern Blot

Rsp frequencies were assessed using Southern blot of restriction endonuclease digested genomic DNA, probed with the H_0 clone. The putative *Rsp* repeats are tandemly arranged, with large intervening blocks of unique sequence DNA. This makes different alleles digested with an appropriate enzyme give a complex banding pattern that is analogous to a DNA "fingerprint".

A suite of enzymes was tested and MspI was adopted as a diagnostic enzyme. When digested with MspI, separated on a Southern blot and probed with labeled H_0 DNA, Kona 13 shows a unique pattern of bands when compared with Kona 3.

The banding pattern is such that the *Kona 13* Rsp^s pattern obscures the *Kona 3* Rsp^i pattern. Therefore, the Rsp^s signature was treated as a dominant marker for frequency analysis.

Heritability of the banding patterns was confirmed by a test cross. F1 offspring from single pair matings between Kona 13 and Kona 3 parents were individually crossed to Kona 13 homozygotes. This was expected to produce 1:1 ratio in the offspring. This was observed (See Figure 4).

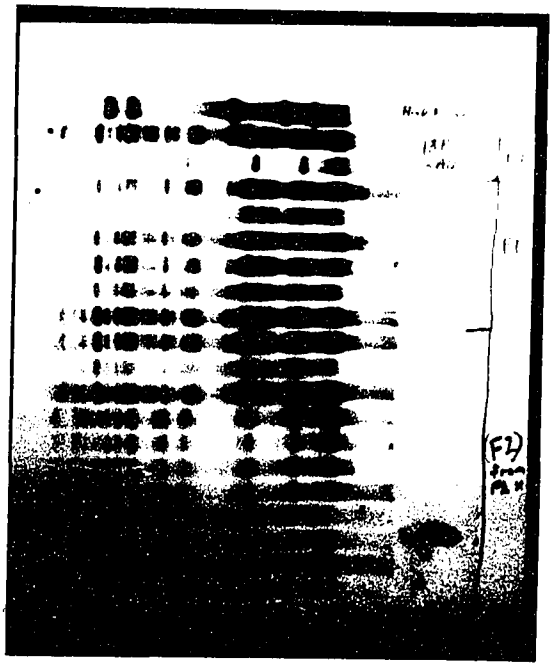


Figure 4. Southern blot showing inheritance of *Rsp* fingerprints. Genomic DNA digested with Msp1 and hybridized with labeled H_0 probe.

Estimates of frequency of insensitive and insensitive alleles were calculated by the Method of Maximum Likelihood (Eliason, 1993, Yasuda and Kimura. 1968). This was necessary for the cages (cages A and D) where *Sd* is present. While the sensitive banding phenotype is dominant to insensitive on the MspI digest blots, the frequency of *Sd*, which is a portion of the apparent insensitive class, is known from the EcoR1 digest blots. Including this information makes a complex formula with no explicit solution for frequency based on counts. The best estimate, therefore, is an MLE. Derivation of the formulae was done using Macsyma (Symbolics, Inc., Cambridge Mass) on a UNIX mainframe at the University of Hawaii Computer Center.

The frequency of Responder sensitive alleles (p) is given by:

$$\text{MLE}(p) = \frac{(\sqrt{N_1 N_2 + N_1^2} + N_1 + N_2)}{(N_1 + N_2)}$$

The variance of the estimate is the negative reciprocal of the second derivative of the generating function. In this case, the variance is:

$$\text{Var} = \frac{(p^6 - 6p^5 + 13p^4 - 12p^3 + 4p^2)}{(2N_1 + 2N_2)p^4 + (-8N_1 - 8N_2)p^3 + (14N_1 + 8N_2)p^2 - 12N_1 p + 4N_1}$$

The frequency of Responder insensitive alleles is then given by:

$$q = 1 - p - r.$$

2.4 Cage Experiments

2.4.1 Standard *D. melanogaster* Stocks

cn bw: This stock carries two second chromosome mutant cinnabar and brown eye color. The cn bw chromosome is an SD+ chromosome that is typically extremely sensitive to SD action, and has been used as the standard sensitive Rsp^s chromosome in many studies.

Cy/cn bw: The S^2 *Cy O cn bw* chromosome carries 5 morphological mutations and multiple inversions. It was used in the standard test for viability and as a balancer chromosome. The stock is in a Π2 cytological background. This is a P cytotype which suppresses P element induced hybrid dysgenesis.

Oregon R balanced stock: The wild-type chromosome, Oregon R, balanced over S^2 *Cy O cn bw*. The S^2 *Cy O cn bw* chromosome has multiple inversions that effectively eliminate recombination.

SD79 bw / S^2 *Cy O cn bw* : *SD79 bw* is the standard lab SD chromosome with a brown eye color marker on the second chromosome. It is maintained balanced over S^2 *Cy O cn bw*. *SD79 bw* has high drive strength (typically k=1.0 against *cn bw*)

2.4.2 Kona *D. melanogaster* Isofemale Stocks

Stocks were derived from wild-caught isofemale lines collected in 1987 in Kona on the Big Island of Hawaii.

Kona 13 (Sd^+ Rsp^s): This isoline was derived from a single wild-caught female and made homozygous for the second chromosome through three generations of sib mating. It was identified as Rsp^s by outcrossing to SD79bw as described below. It was further checked via Southern blot for *Rsp* repeat structure content. It was maintained as a dump stock.

Kona 3 (Sd^+ Rsp^i): This isoline was derived from a single wild-caught female and made homozygous for the second chromosome through three generations of sib mating. It was identified as Rsp^i by outcrossing to SD79bw as described below. It was further checked via Southern blot for *Rsp* repeat structure content. It was maintained as a dump stock.

Kona 7 (Sd^+ Rsp^i): This isoline was derived from a single wild-caught female. It carries a unique second chromosome inversion (55c to 47e-f). It was used as a balancer chromosome for *SD10*.

Kona *SD10* (*Sd* Rsp^i): This second chromosome was initially identified as SD cytologically by the presence of In(2R)NS inversion at 52A to 56F and the small pericentric inversion (39-40 to 42A) typical of SD-72 type chromosomes. Drive strength was tested by outcrossing to tester chromosomes as described below. The

chromosome is maintained as a cage stock, balanced over Kona 7 *(Sd^+ Rsp^i)*:

Kona *SD10bw (Sd Rsp^i bw)*: This chromosome is a derivative of Kona *SD10* carrying a brown eye color marker on the second chromosome. It was derived by recombination with *cnbw*. It carries In(2R)NS and has full drive strength. Maintained balanced over S^2*CyOcnbw* in a Π^2 background.

2.4.3 Establishment of Kona *SD10*/ Kona 7 Cage Stock

Because Kona *SD10* proved to be homozygous infertile in males (see Results section), it was necessary to maintain the stock with a balancer chromosome. Another isoline from the same collection, Kona 7, was found to have a unique inversion that overlaps the In(2R)NS inversion in 2R that is associated with SD-72 types. Kona *SD10* carries *Sd* with the In(2R)NS inversion and the pericentric inversion. The In(2R)NS inversion is at 52a to 56f and the #7 inversion is a 55c to 47e-f. The overlapping inversions should suppress recombination in the region, so that the *SD10* chromosome remains intact. This balanced stock, therefore, contain only chromosomes from the same natural population.

Flies from the #10 isoline were crossed with flies from the # 7 isoline in 30 single pair matings. Several larvae from each cross were checked cytologically until one was found to have both inversions. Offspring from that line were mated and F_2 larvae were checked

until a line was found that appeared to be from 2 heterozygous parents. Offspring from these were mated and lines F3 larvae were checked to see if the heterozygotes were persisting. At that point, the line was expanded into a bottle and then after a few generations, transferred to a population cage. This cage stock was maintained and checked occasionally via cytology for the presence of both inversion haplotypes and via Southern blot analysis for *Sd* genotypes.

Drive strength was measured as the *k* value, defined as the proportion of SD-bearing offspring produced by an SD/SD+ heterozygous father. Tests of drive strength for this study were conducted in two ways. The Kona *SD10bw* chromosome is derived from the Kona 10 isoline and a carries a mutant brown allele on the tip of 2R, derived by recombination from the *cnbw/cnbw* stock. It continues to drive strongly against *cnbw*. This phenotypic marker makes it possible to test the chromosome by individually mating *SD10bw/+* fathers (where + is the chromosome being tested) to *cnbw* homozygous females and counting the *bw* vs. wild-type offspring. *SD79bw*, a standard lab SD chromosome) was similarly tested against *Sd+* chromosomes of interest to establish relative sensitivity. Chromosomes with no visible mutations were tested by crossing appropriate parents and identifying the *Sd* genotype of the offspring using the molecular probe PGN156 (described below). Kona *SD10* was tested against *cnbw* Rsp^s, *Kona 13* Rsp^s, *Kona 3* Rsp^i, and *Kona 7* Rsp^s.

Homozygote viability of the chromosomes was assessed using a standard Curly viability test (Mukai, 1964) in which homozygotes are

crossed to a *Cy / cn bw* stock and then Curly winged flies (*Cy / +*) are sib mated and the offspring counted. The *Cy / Cy* homozygotes are inviable and the resulting straight-winged homozygotes and Curly-winged heterozygotes should be in a 1:2 ratio if the homozygotes are fully viable. A reduction from 33% homozygotes represents a viability effect. All were placed as single pairs on food cooked on the same day. The heterozygous crosses were brooded on day three. These were held in a 25° incubator and offspring counted at 14 and 17 days after the original mating. The control was the wild-type chromosome, Oregon R.

2.4.4 Experimental Cages

There were 3 different sets of experimental cages, with 3 replicates each, for a total of 9 cages. The three sets (A, D, and F) differed in their intended starting frequencies. The replicates for each were denoted as the following:

Cage A: A1, A2, A3

Cage D: D1, D2, D3

Cage F: F1, F2, F3

The cages were started with the appropriate numbers of vials from the *Kona 13 Rsp^s* stock, the *Kona 3 Rsp^i* stock and flies made heterozygous for *Kona SD10* and *Kona 3 Rsp^i*. The latter were produced by mating single males from the Kona 7/10 population cage with virgin females from either of the Kona 3 sublines. After a few days, the males were tested by Southern blot analysis for *Sd*

genotype. Those crosses that had *SD10*/iso7 fathers were used for the population cages.

The A cages were each started with 4 Kona 3 vials, 4 Kona 13 vials and 2 of the (*SD10*/7 x Kona 3) vials. This should have given starting frequencies of Rsp^i = 0.45, Rsp^s = 0.45 and *Sd* = 0.1.

The D cages were each started with 1 Kona 3 vial, 7 Kona 13 vials and 2 of the (*SD10*/7 x Kona 3) vials. This should have given starting frequencies of Rsp^i = 0.1, Rsp^s = 0.8 and *Sd* = 0.1.

The F cages were each started with 4 Kona 3 vials and 4 Kona13 vials. This should have given starting frequencies of Rsp^i = 0.5, Rsp^s = 0.5.

The population cages used were standard 28 cm x 15 cm x 15 cm Plexiglas cages. They were kept in a 25° C incubator for the duration of the study. Each cage has openings for 13 food vials. The food was changed on a rotating schedule so that each vial remained in the cage approximately 14 to 20 days.

All cages were started on October 15, 1994. The first collection was made on November 11, 1994. Collections of approximately 60 flies per cage were been made at 2 to 4 week intervals by removing vials from the cage and collecting the adults that hatched over 1 week. Samples were placed in a 1.5 ml microcentrifuge tube, labeled and frozen at -70° C for later use.

Chapter 3
Results

3.1 Kona Chromosomes Analysis

Four isofemale lines were examined: *Kona SD10*, *Kona 3* Rsp^i, and *Kona 13* Rsp^s using molecular, cytological, and genetic means. Results of k tests are summarized in Table 1

Table 1: Tests of drive strength of various SD chromosomes against *Rsp* where k is the drive strength, σ is the standard deviation of the samples and N is the total number of offspring scored.

SD	target	k	σ	N	date
SD79bw	Kona 3	0.520	0.021	2272	10/2/94
SD79bw	Kona 3	0.545	0.095	415	10/2/94
SD79bw	Kona 13	0.848	0.052	1314	10/2/94
SD79bw	cnbw	0.996	0.007	531	10/2/94
SD79bw	Kona 3	0.550	0.025	1665	10/8/93
SD79bw	Kona 3	0.490	0.040	330	10/8/93
SD79bw	Kona 13	0.764	0.076	481	10/8/93
SD79bw	cnbw	0.994	0.009	537	10/8/93
SD10bw	Kona 3	0.461	0.077	701	1/24/93
SD10bw	Kona 13	0.941	0.055	633	1/24/93
SD10bw	cnbw	0.982	0.027	224	1/24/93
SD10bw	Kona 7	0.930	0.061	631	4/15/89
SD10bw	cnbw	1.0	0.0	504	4/15/89

3.1.1 Molecular and Genetic Analysis of *SD10*

Southern blots of genomic digests with EcoR1, hybridized with labeled PGN156 (*SD*) probe show a single 12 Kb band in *Kona SD10* homozygotes corresponding to the *Sd* allele. *SD10/Sd*$^+$ heterozygotes show two bands of 12 and 7 Kb and *Sd*$^+$*/Sd*$^+$ homozygotes show a single band of 7 Kb.

Standard tests of sensitivity of Kona *SD10* were done using the recombinant chromosome *SD10bw*. The results (Table 1) show that *SD10bw* has high drive against sensitive responders.

It was impractical to do a standard test of drive strength of *SD10* chromosome in its native background with Kona 7. The molecular visualization technique would have required a very large number of Southern blots for the test. However, a rough equivalent was done using mass DNA preparations (approximately 100 flies each) from F1 flies derived from of crosses between *SD10/Kona7(Rsp*^{s}*)* males and *Kona 13 Rsp*^{i} females. The *Kona 7 Rsp*^{s} chromosomes do not appear in the offspring, indicating that these fathers apparently produce no (or very few) Kona 7-bearing sperm. This is based on 23 individual crosses and suggests very strong drive strength.

SD10bw was also used to estimate viability of *SD10bw* in a standard Curly viability test. The viability of *SD10* homozygotes in this test was 0.298 (σ=0.0716, N=684). This is a fitness of 0.894

(0.298/0.33) relative to Curly. A control using Oregon R (a standard lab wild-type line), gave a viability of 0.34 (σ=0.0837, N=763).

In single pair crosses of males from the *SD10*/iso7 cage, 23% (45/195) of the crosses were non-productive. Most of the non-productive fathers (those that survived after one week) were checked by Southern blot analysis and found to be *SD10/SD10*. The frequency of *Sd* in this cage at that time was 0.57, estimated from Southern blot analysis of 38 individuals. This frequency gives a Hardy-Weinberg expectation of 32% SD homozygotes. A simple goodness of fit test of the observed 23% with this expectation rejects the null hypothesis of Hardy-Weinberg expectation (χ^2 = 7.14, with 1 degree of freedom). The deviation is most likely the result of gametic disequilibrium resulting from meiotic drive in males, which would tend to reduce the number of homozygotes observed. However, it may also represent a slight reduction in viability of the SD homozygotes.

3.1.2 Molecular and Genetic Analysis of Kona 3 *Rsp^i*

Partial digests with XbaI and P2.5 probe show approximately 18 bands (Figure 5) which identifies these as *Rsp* insensitive. This is supported by dot blot data, with none of five flies tested hybridizing to the probe. MspI digests produced a few bands of different sizes, showing that the few repeats present are interspersed with large blocks of non-repetitive DNA differing either in length or MspI

restriction sites. Probing with PGN156 (*Sd* probe) shows that 1) only one band, corresponding to the Sd^+ allele, is produced in an EcoR1 digest, and 2) MspI digests fail to distinguish *Sd* from Sd^+.

Four separate tests of sensitivity of Kona 3 chromosomes to drive by SD79bw were performed in 1993 and 1994 (Table 1). On this basis, it was formally designated Rsp^s.

Viability of Kona 3 was estimated in a standard Curly viability test. The viability of Kona 3 homozygotes in this test was 0.3070 (σ=0.0616, N=811). This is a fitness of 0.922 (0.307/0.33) relative to Curly. A control using Oregon R (a standard lab wild-type line), gave a viability of 0.34 (σ=0.0837, N=763).

3.1.3 Molecular and Genetic Analysis of 13 Rsp^s

Partial digests with XbaI P2.5 probe show >>100 bands (Figure 5) which identifies it as *Rsp* sensitive. This is also supported by dot blot data, with five of five flies tested hybridizing very strongly to the probe. Furthermore, MspI digests produce a distinctive pattern of bands (compared with other presumptive chromosomes from the same population including Kona iso7 (discussed above) and with cnbw) showing that the repeats are interspersed with non-repetitive DNA differing either in length or MspI restriction sites. Probing with PGN156 (*Sd* probe) shows that 1) only one band, corresponding to the Sd^+ allele, is produced in an EcoR1 digest, and 2) MspI digests fail to distinguish *Sd* from Sd^+.

Two separate tests of sensitivity of Kona 13 to drive by SD79bw were performed in 1993 and 1994 (Table 1). On this basis, it was formally designated *Rsp*^{i}.

Viability of Kona 13 was estimated in a standard Curly viability test. The viability of Kona 13 homozygotes in this test was 0.331 (σ=0.0811, N=857). This is a fitness of 0.99 (0.331/0.33) relative to Curly. A control using Oregon R (a standard lab wild-type line), gave a viability of 0.34 (σ=0.0837, N=763).

3.2 SD Banding Pattern

A typical blot is shown in Figure 6 illustrating the diagnostic banding patterns that were used to distinguish between SD homozygotes and heterozygotes. Homozygotes show a single band of either 7 Kb (*Sd*$^{+}$) or a 12 Kb (*Sd*), while heterozygotes show both bands. Thus, the homozygote and heterozygote frequencies were estimated by direct count of alleles.

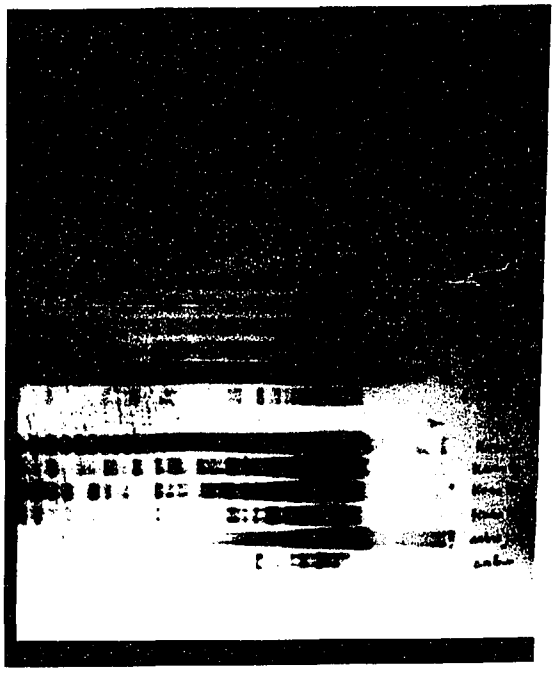


Figure 5. Southern blot showing *Rsp* copy number in Kona isolines. Genomic DNA partially digested with Xba1 and hybridized with labeled PGN156 probe.

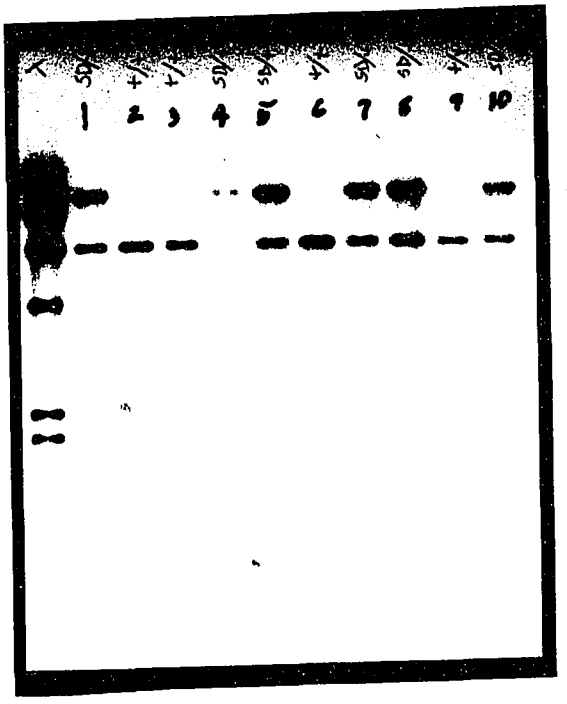


Figure 6. Typical blot showing SD banding. Genomic DNA digested with EcoR1 and hybridized with labeled PGN156 probe.

3.3 *Rsp* Banding Fingerprints

A typical blot is shown in Figure 7 illustrating the diagnostic banding patterns that were used to distinguish between Rsp^i homozygotes and Rsp^s homozygotes and heterozygotes. Note that an individual heterozygous for *Kona SD10* and either *Kona13* or *Kona 3* could not be distinguished from the *Kona13* or *Kona 3* homozygotes respectively.

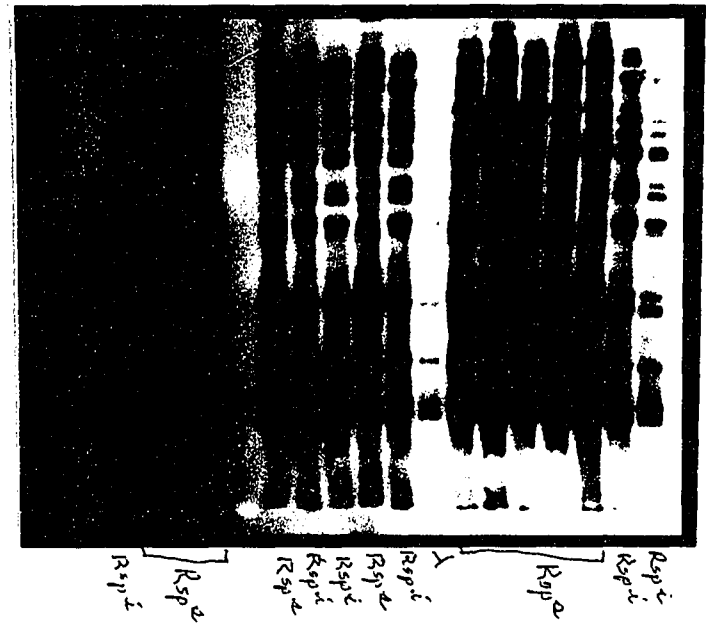


Figure 7. Typical blot showing *Rsp* fingerprints. Genomic DNA digested with Msp1 and hybridized with labeled H_0 probe.

3.4 Simulation Model Results

Population dynamics simulations of meiotic drive versus directed mutation of 4 basic types have been produced. Some general comments apply to all the models:

In the ranges of parameter values examined, all models produce stable equilibria for wide ranges of parameters but some combinations result in loss of one or the other components.

All the models are robust with respect to starting frequencies, i.e., a given set of parameters result in the same final values regardless of what the starting frequencies were.

For relatively high values of meiotic drive (k>0.6) and mutational event rate (PU>0.005), stability is reached in 15-50 generations. For lower values, the times required are greater but usually not more than a few hundred generations, except for very extreme values.

3.4.1 USCE

The unequal sister chromatid exchange model predicts stable equilibria with moderately low frequency of *Sd* (less than 0.2) but low frequency of sensitive Responder classes (less than 0.1 each). The truncation selection version and the bounceback versions produce the same basic patterns, with the *Sd* frequency being very slightly higher relative to insensitive with truncation selection (Figures 8 and 10). Equilibrium frequency of Responder is dependent

on both PU and k, with moderate changes in frequency over wide ranges of k and somewhat more response to PU (Figures 9a and 11a). Equilibrium frequency of *Sd* is mainly dependent on PU, with k being less important (Figures 9b and 11b).

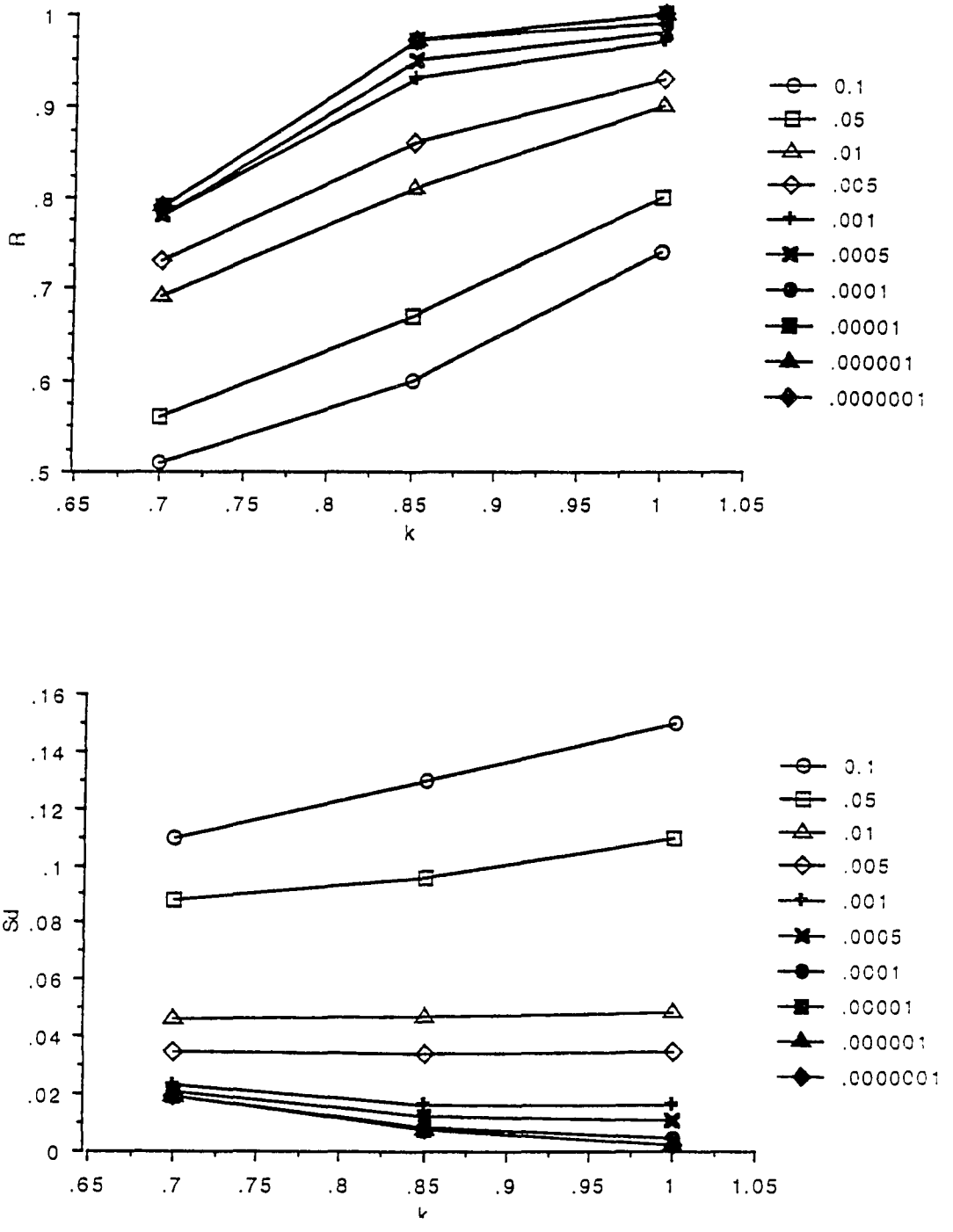


Figure 8. USCE Bounceback model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

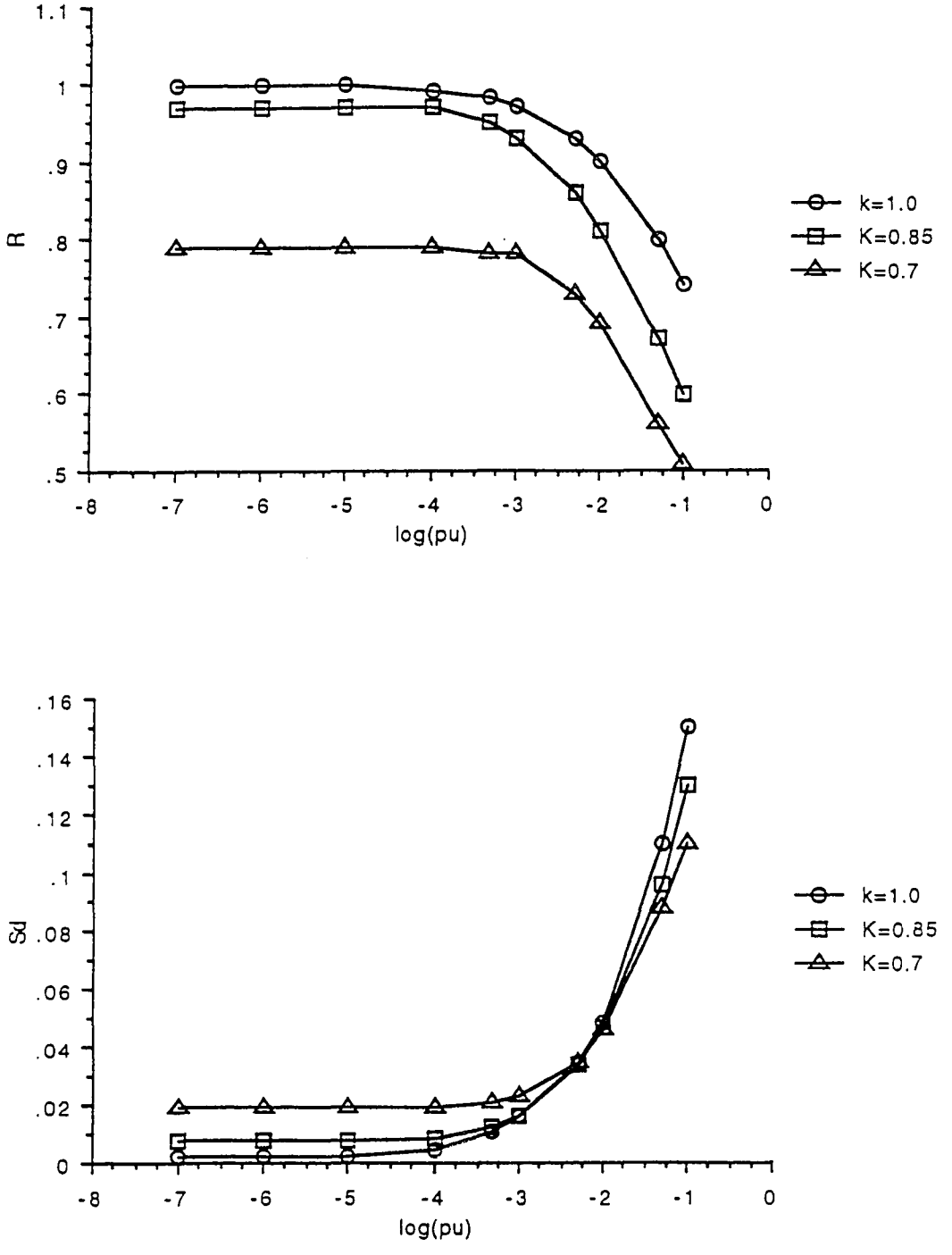


Figure 9. USCE Bounceback model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

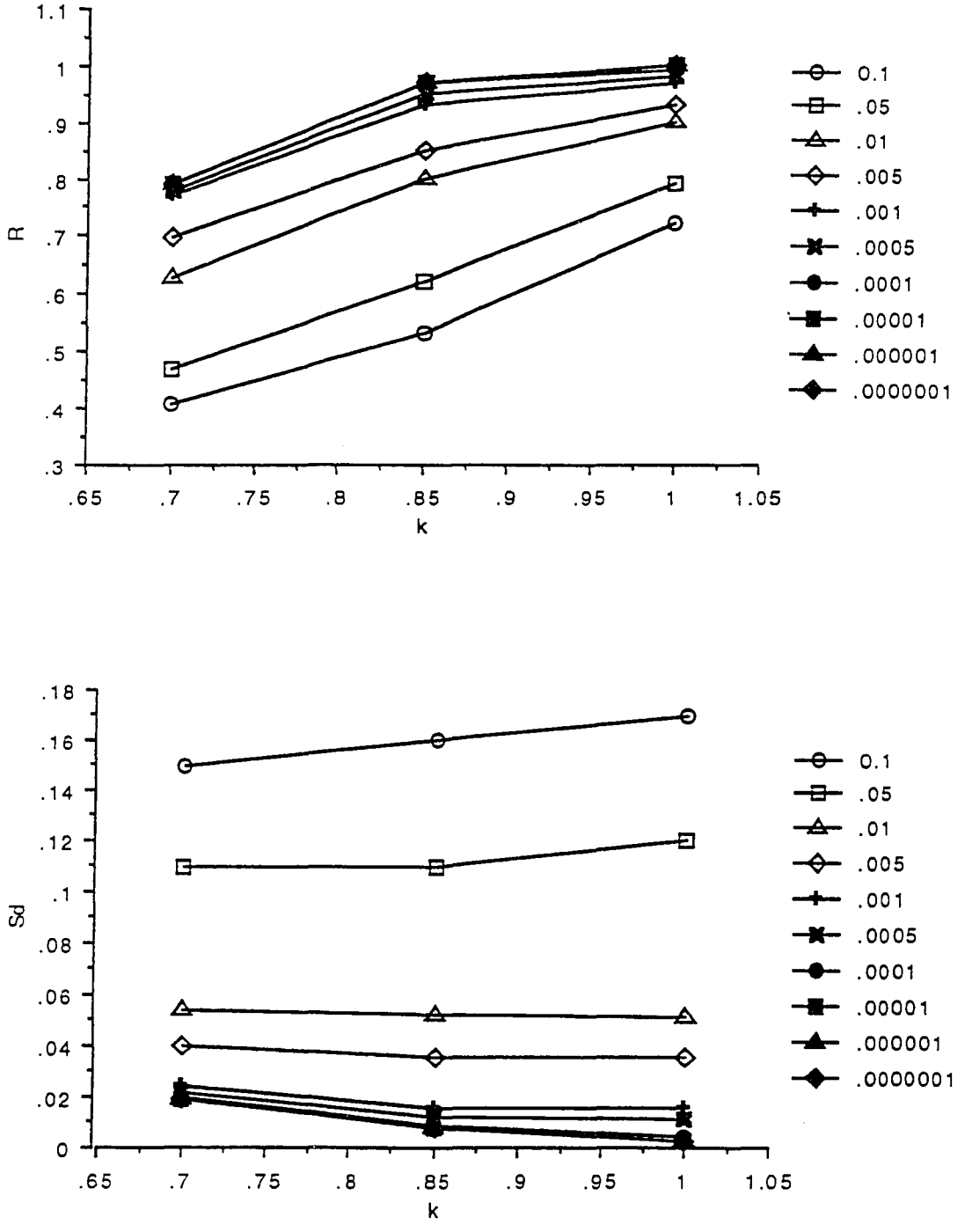


Figure 10. USCE Truncation model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

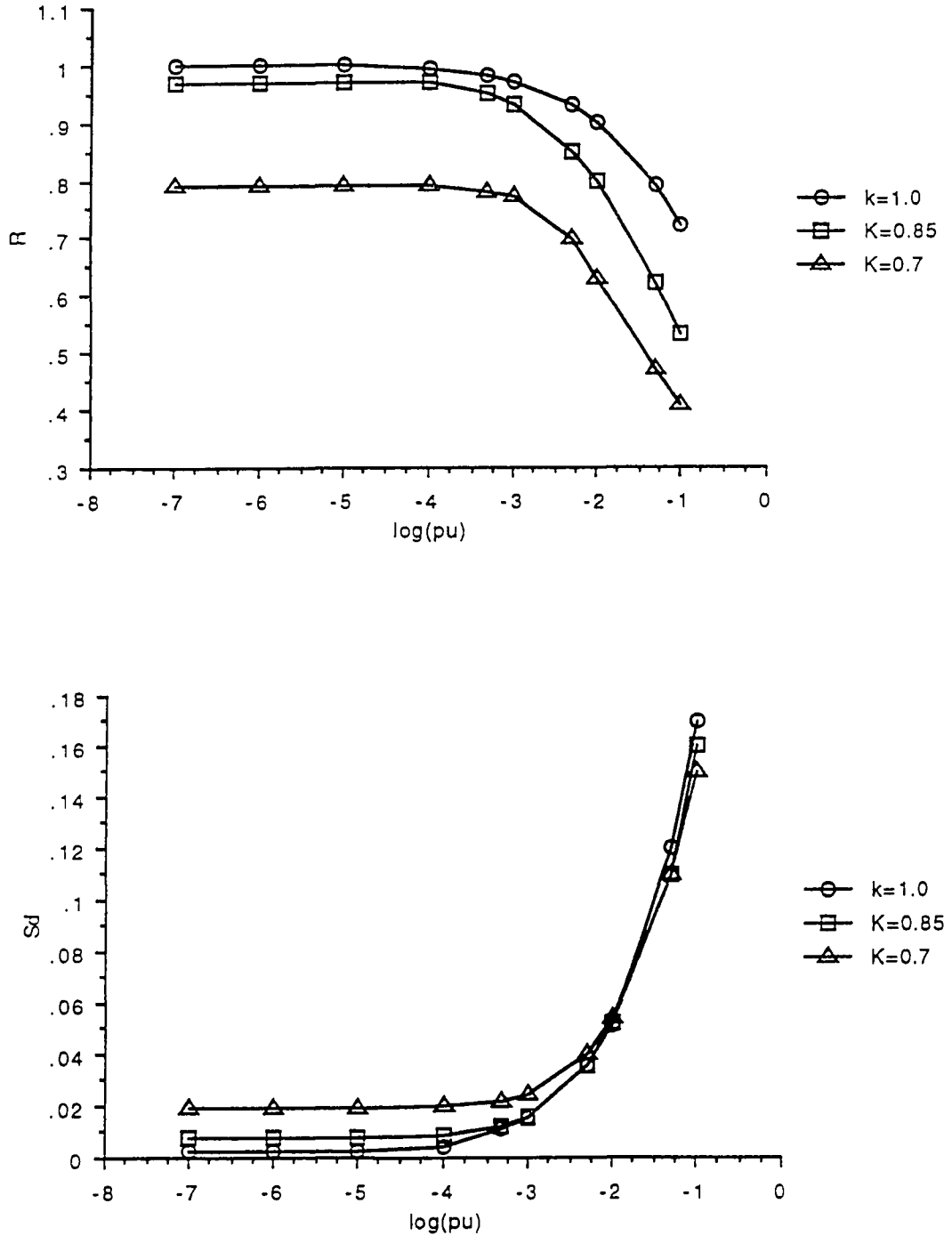


Figure 11. USCE Truncation model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

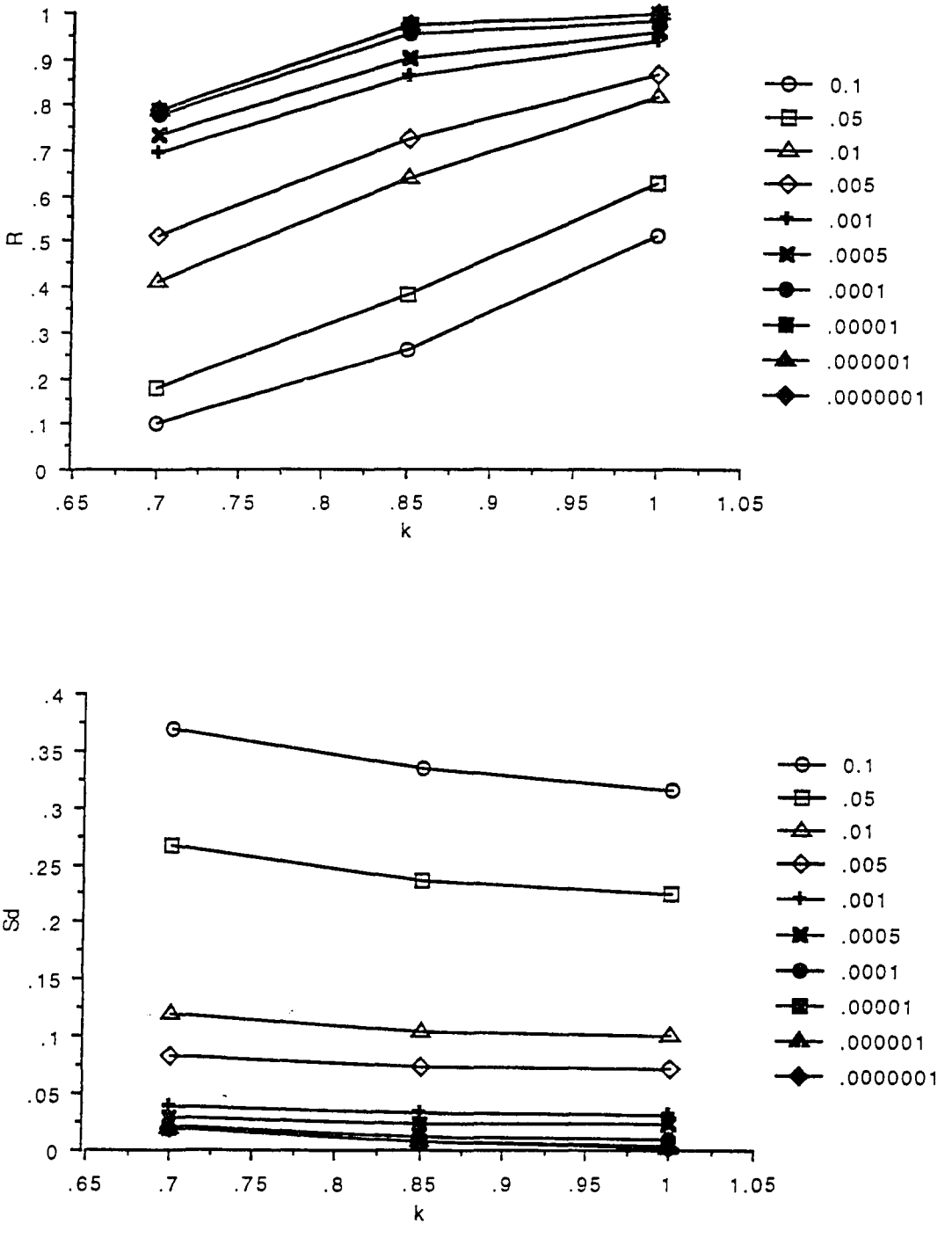


Figure 12. DSBR model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

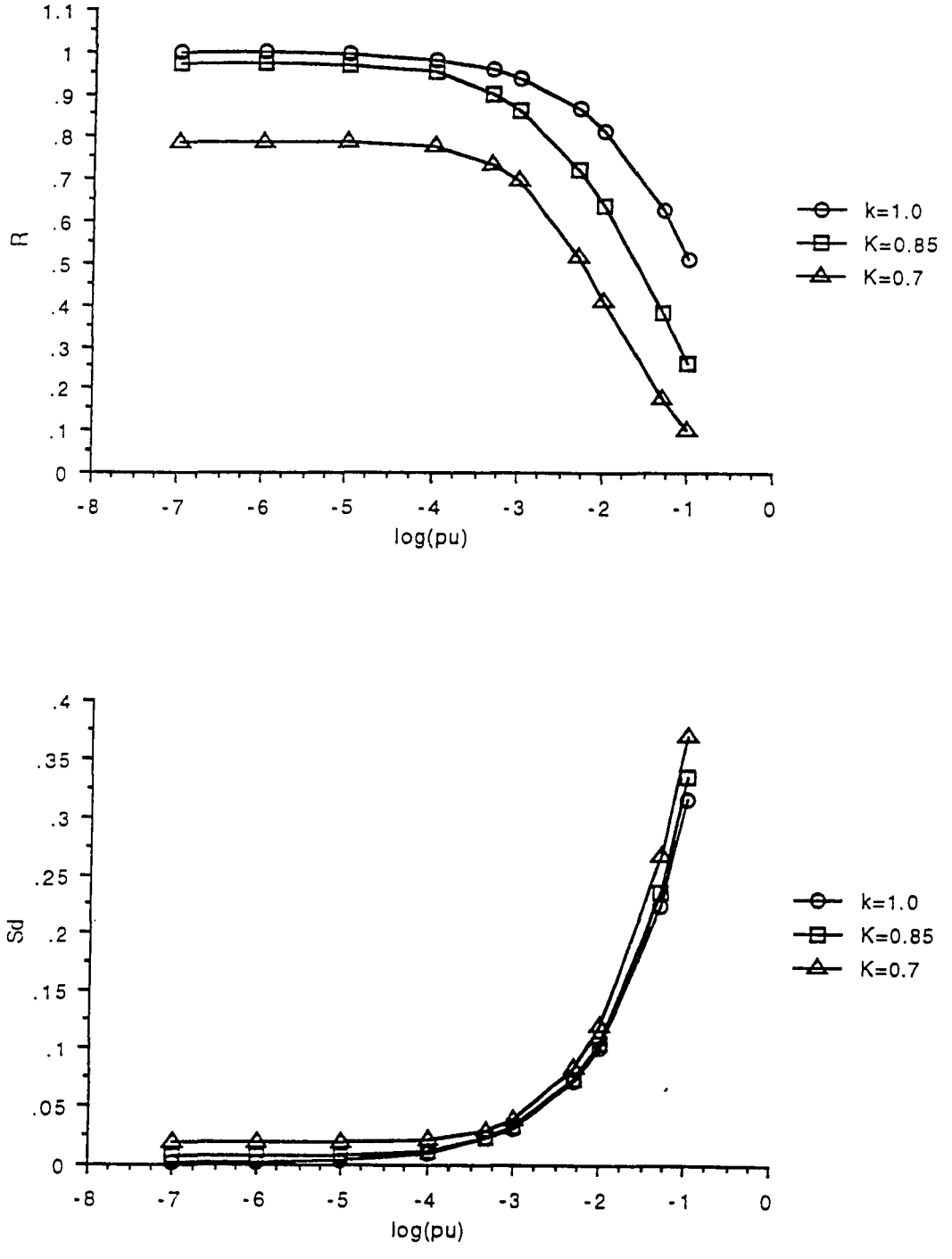


Figure 13. DSBR model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

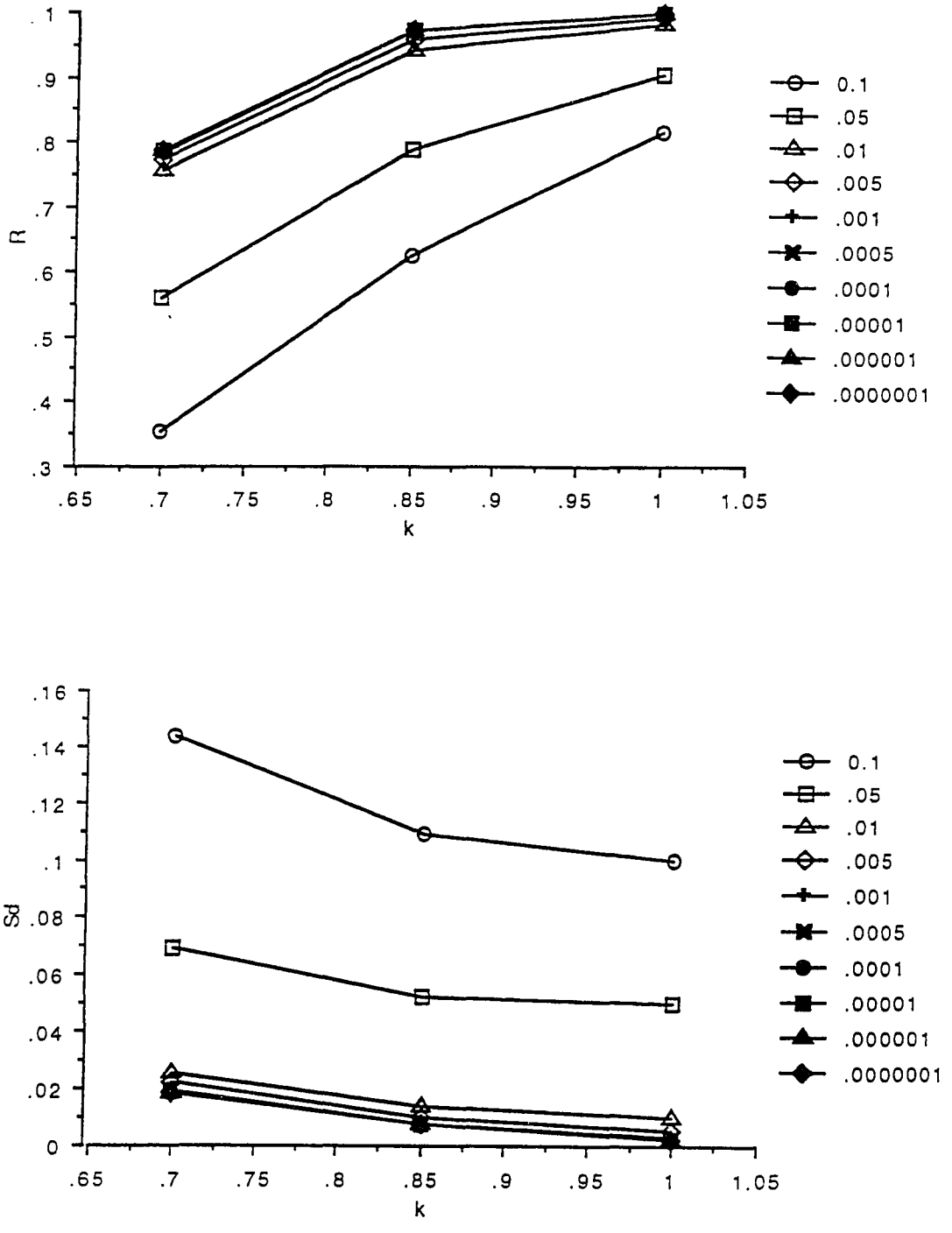


Figure 14. Hits model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

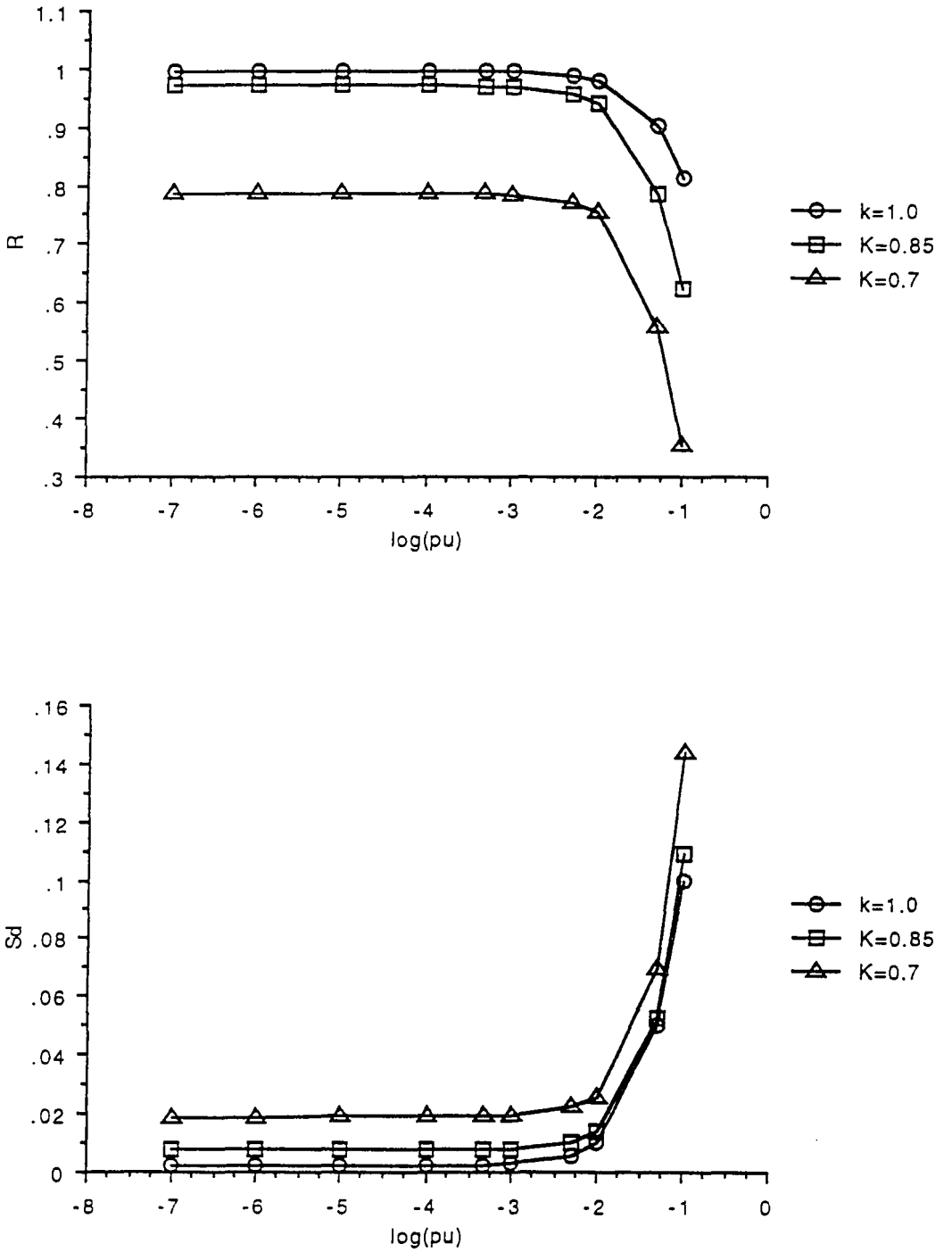


Figure 15. Hits model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

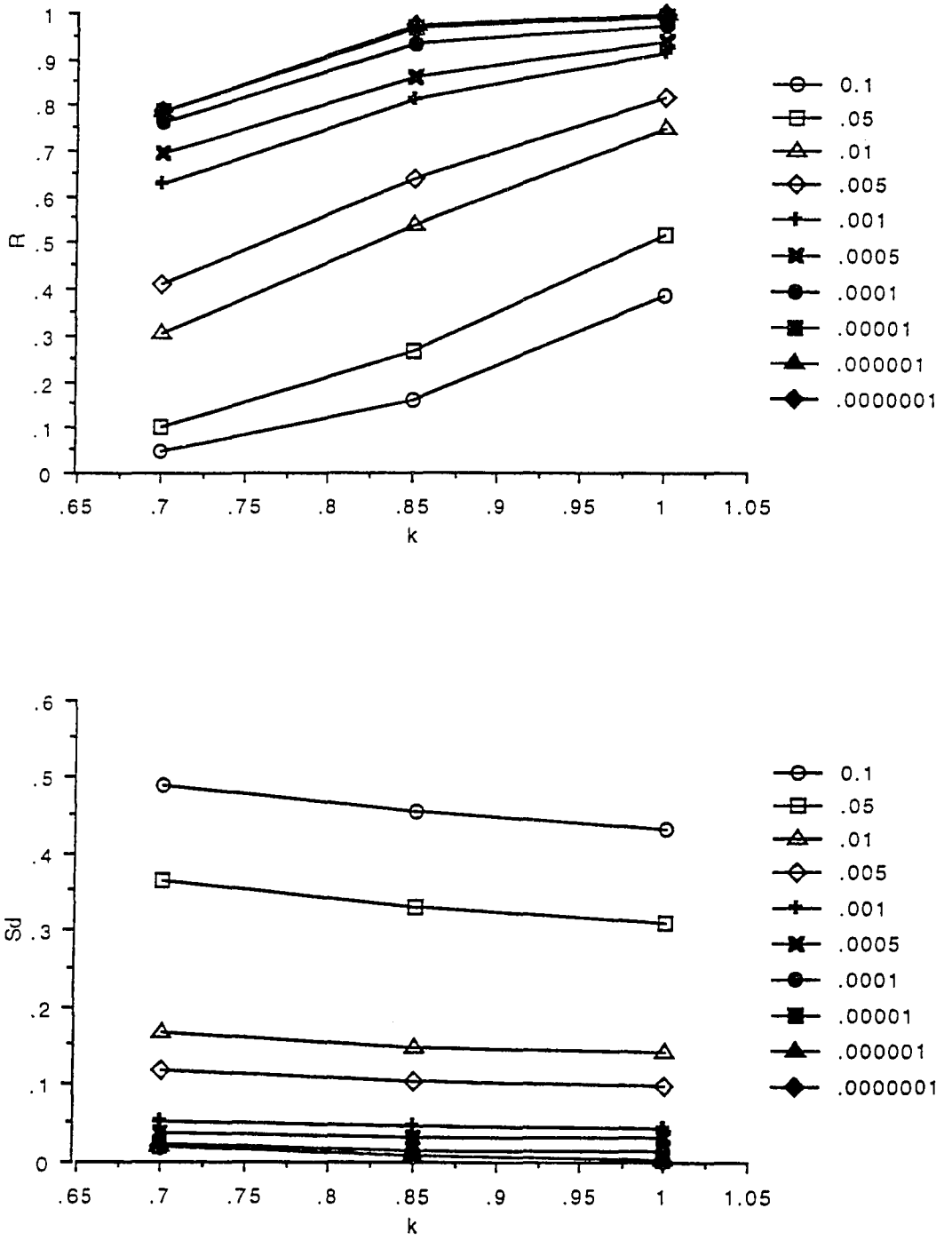


Figure 16. RSRamm model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

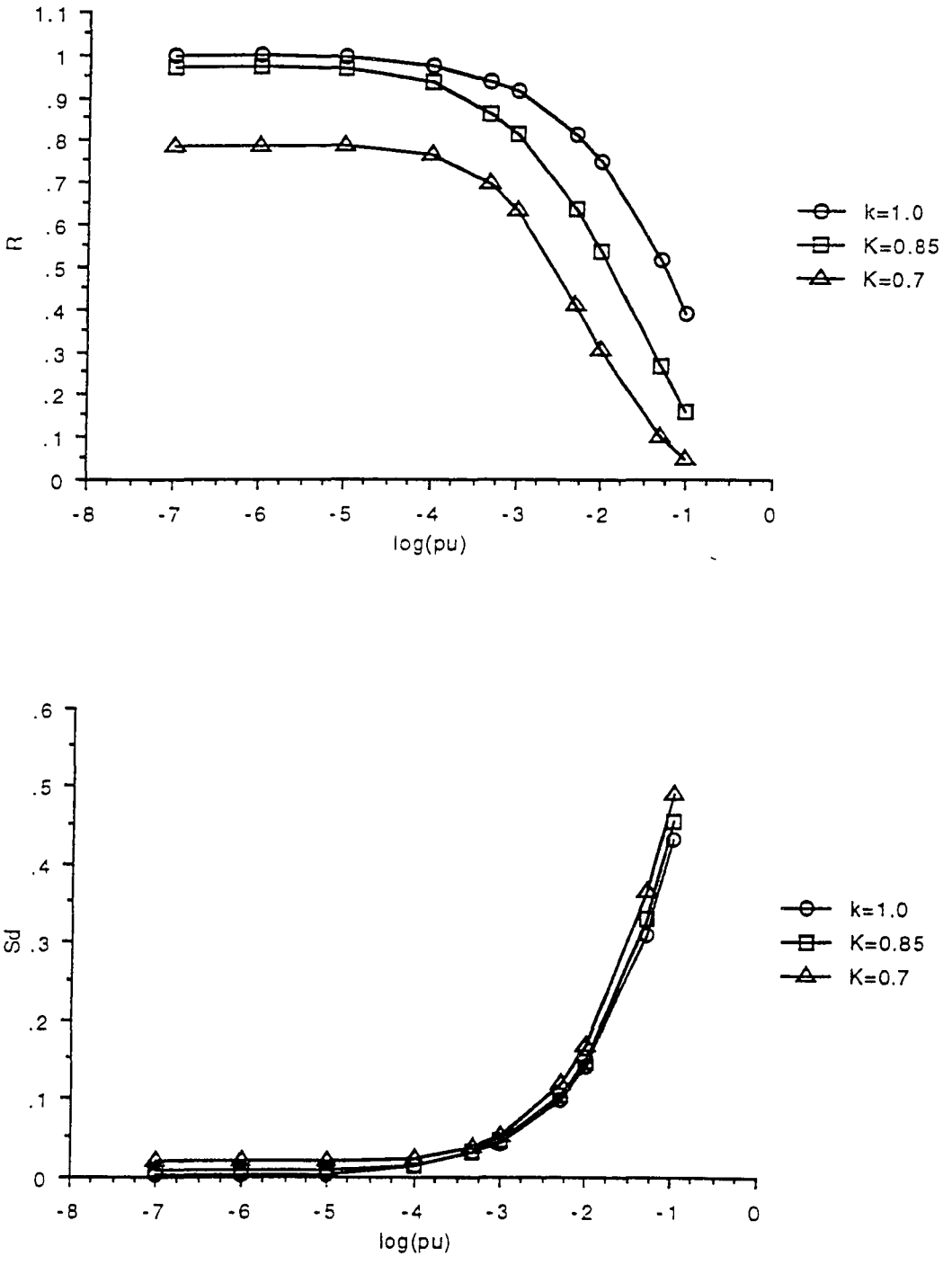


Figure 17. RSRamm model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

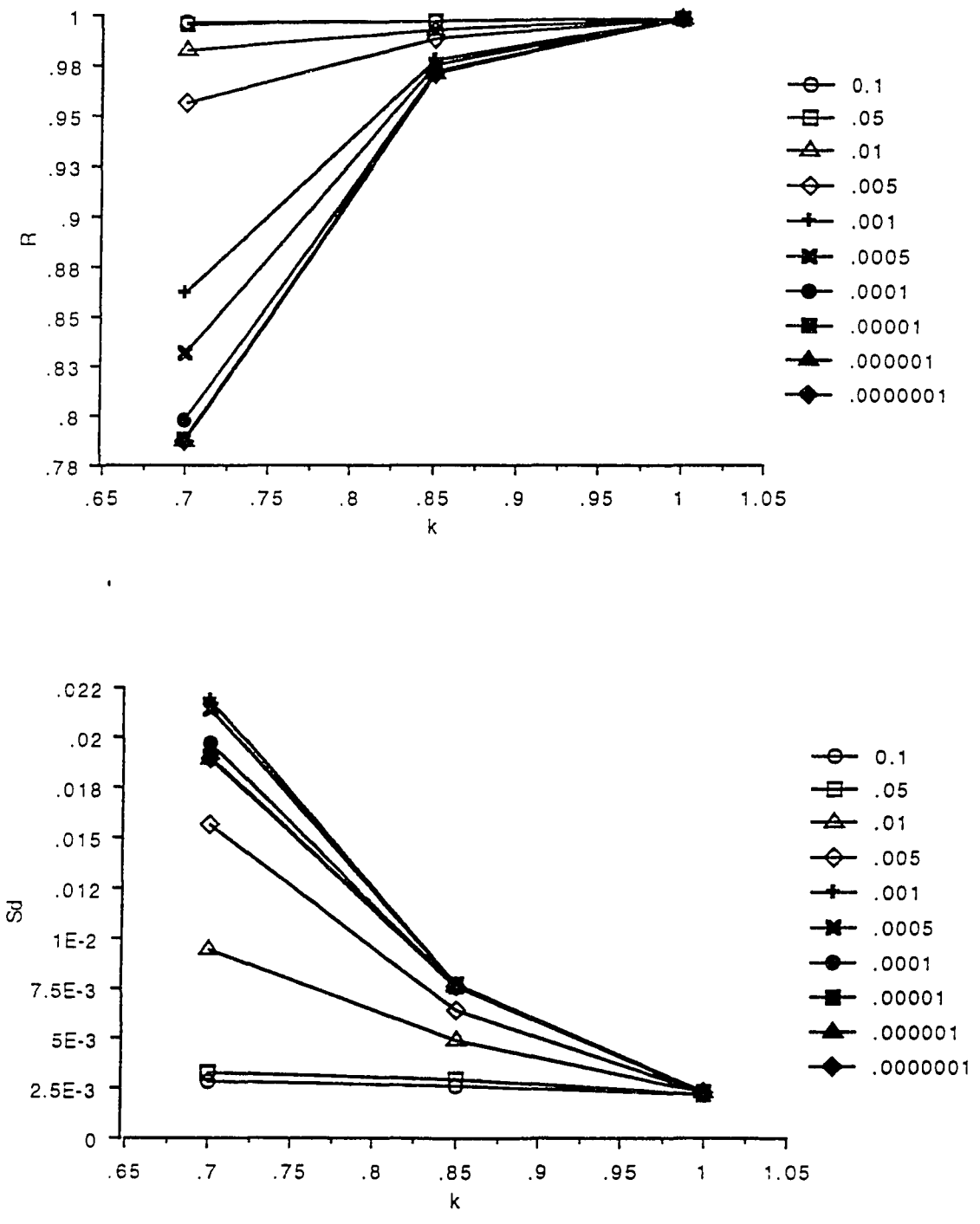


Figure 18. RSRsrm model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus k values for 10 different values of PU.

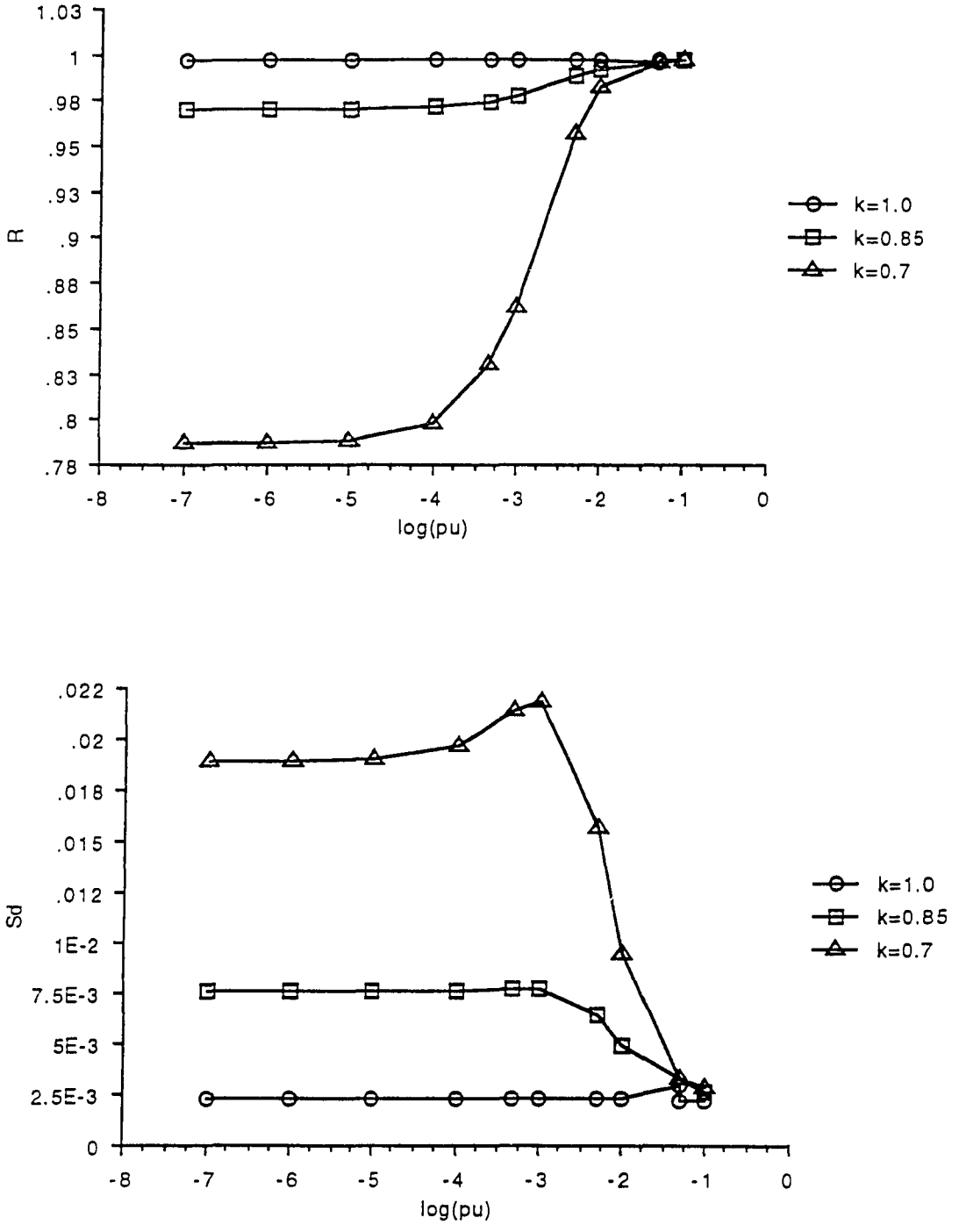


Figure 19. RSRsrm model: Equilibrium frequencies for (a) *Rsp* and (b) *Sd* versus log(PU) for 3 different k values.

3.4.2 DSBR

The double strand break repair model predicts equilibria with higher levels of sensitive Responder but also higher frequencies of *Sd* (Fig 12). Equilibrium frequency of Responder is dependent on both PU and k (Fig 13a). Equilibrium frequency of *Sd* is mainly dependent on PU, with k being less important (Figs 13b).

3.4.3 Hits

The *Sd* moderated model predicts very high levels of *Rsp* insensitive except for the highest PU levels examined and low drive levels (Figs 14a and 15a). *Sd* frequencies are correspondingly low, except for the case of high PU (Figs 14b and 15b).

3.4.4 RSR

The two different rules on misalignment size produce very different results. The allele maximum misalignment rule (amm) produces equilibrium values very similar to those of the USCE bounceback model, with slightly higher *Sd* frequencies (Figs 16 and 17). On the other hand, the single repeat misalignment rule (srm) leads to fixation of insensitive except for the case of low k (k=0.7) in which final value of insensitive is highly dependent on PU (Figs 18a

and 19a). Similarly, *Sd* is at very low frequencies regardless of PU expect for the case of k=0.7 (Figs 18b and 19b).

3.5 Experimental Cage Results for Cages A and D

Among the six cages (cages A1, A2, A3, D1, D2, and D3) containing *Sd*, a total of 18 samples were assessed for *Rsp* and *Sd* frequencies using Southern blot analysis. Results are summarized in Table 2.

Table 2: *Sd* frequencies are in the form: Freq. (sample size). Responder sensitive frequencies are given in the form: Freq. ± Var (sample size). The variance is the variance of the estimate of frequency from the MLE. Sample sizes are number of scorable lanes for the *Sd* and *Rsp* Southern blots, respectively.

Cage	allele	11/18/94	2/3/95	9/7/95
A1	*Sd*	0.190 (18)	0.36 (18)	0.25 (16)
	sens	0.156±0.0037 (20)	0.20±0.0048 (19)	0 ±0.0 (19)
	ins	0.654	0.44	0.75
A2	*Sd*	0.1969 (33)	0.429 (14)	0.156 (16)
	sens	0.169±0.0018 (42)	0.118±0.0031 (18)	0 ±0.0 (20)
	ins	0.6341	0.453	0.844
A3	*Sd*	0.2 (20)	0.236 (19)	0.325 (20)
	sens	0.134±0.0031 (20)	0.142±0.0035 (19)	0 ±0.0 (30)
	ins	0.666	0.622	0.675
D1	*Sd*	0.183 (30)	0.196	0.263 (19)
	sens	0.042±0.0017 (12)	0 ±0.0 (30)	0 ±0.0 (30)
	ins	0.775	0.804	0.737
D2	*Sd*	0.333 (12)	0.361 (18)	0.375 (20)
	sens	0.2697±0.0078 (15)	0.1835±0.0046 (18)	0 ±0.0 (30)
	ins	0.3973	0.4555	0.625
D3	*Sd*	0.361 (18)	0.294 (17)	0.4 (20)
	sens	0.0814±0.0012 (32)	0.0267±0.0001 (19)	0 ±0.0 (30)
	ins	0.5576	0.6793	0.6

3.6 χ^2 Comparison of Cages A and D with Simulation Model Results

The goodness of fit of the observed cage frequencies for those cages that had *Sd* (cage sets A and D) was assessed using a minimum χ^2 approach. A series of iterations of each simulation was used to generate expected frequencies at generation times corresponding to those of the actual collections. The null hypothesis is that the observed values do not differ from the expected values. The observed frequencies of each cage were compared against the null hypothesis generated by each model over a range of k and PU values. The combination producing the minimum χ^2 value is thus the best fit to the model.

Since the *Sd* frequencies and the *Rsp* frequencies were measured on separate sets of flies they are best considered separate tests. Furthermore, since the molecular phenotypes produce results that overlap across the 2 samples (i.e., in the *Sd* test, non-*Sd* contains both Rsp^s and Rsp^i and in the *Rsp* test, Rsp^s is dominant to both Rsp^i and *Sd*), the two tests samples are not fully independent. For these reasons, it is necessary to assess the minimum χ^2 analyses separately on *Sd* data and *Rsp* data.

An additional complication was introduced by the fact that the starting cage frequencies turned out to be quite different in certain cases. In particular, although the A cages had similar starting frequencies, the D cages had starting frequencies that were very

different from one another and from what was intended. This is probably best explained by the fluctuations in population composition produced by the expansion of the population from a few hundred to thousands in the cage. It has been observed (Lyttle, pers. comm.) that in the early generations, allele frequencies in new cages can fluctuate considerably and it is often difficult to hit intended starting frequencies. In this case, cages D1 and D3 had extremely low frequency of Rsp^s in the early generations. Whatever the reason, the result of this is that the cages cannot properly be considered replicates of the same starting frequencies but rather are separate tests that may have different starting frequencies. For this reason, the minimum χ^2 analyses were carried out on each cage independently. It turned out that, in general, the three A cages gave results that were very similar while the three D cages varied more.

An overall best joint fit was identified by summing χ^2 values for *Sd* and *Rsp* frequencies for each set of parameters. However, this value is suspect statistically due to the non independence of the observations. Therefore, the result is only qualitative, not quantitative.

None of the six models tested gives a χ^2 small enough to be considered a good fit for both variables, as discussed below. In many cases, the range of PU values that gave a very small χ^2 for one set of frequencies gave produced very large χ^2 for the other.

3.6.1 χ^2 Analysis of the USCE Model

The unequal sister chromatid exchange model with truncation selection (Figures 20a-f and Table 3) and the model with bounceback (Figures 21a-f and Table 4) produce the same basic

Table 3a. Minimum χ^2 for the fit of the observed values from cages A and D to the USCE Truncation Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.34	4.27	0.99	7.11
	0.85	0.67	6.87	0.10	16.21
	0.7	0.99	4.99	0.10	24.94
A2	1	0.32	7.7	0.10	5.24
	0.85	0.66	9.51	0.10	18.54
	0.7	0.99	9.18	0.10	28.94
A3	1	0.32	4.71	0.99	9.49
	0.85	0.57	4.37	0.10	25.94
	0.7	0.90	4.42	0.10	39.43
D1	1	0.42	7.36	0.99	11.97
	0.85	0.81	7.84	0.10	31.23
	0.7	0.99	7.75	0.10	42.35
D2	1	0.57	15.3	0.10	9.43
	0.85	0.99	15.36	0.10	26.26
	0.7	0.99	17.58	0.10	40.27
D3	1	0.48	8.74	0.10	9.34
	0.85	0.87	8.64	0.10	29.52
	0.7	0.99	9.47	0.10	44.64

Table 3b. Joint best fit χ^2 for Cages A and D to the USCE Truncation Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	1	0.34	0.13	7.27
A2	1	0.28	0.74	7.44
A3	1	0.28	0.86	10.6
D1	1	0.42	0.36	12.02
D2	1	0.6	5.34	11.63
D3	1	0.45	0.78	16.24

Table 3c. Joint best fit χ^2 for Cages A and D to the USCE Truncation Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.99	6.48	29.85
A2	0.1	18.83	17.57
A3	0.1	19.01	25.7
D1	0.99	8.13	41.44
D2	0.99	15.36	44.1
D3	0.19	24.58	46.96

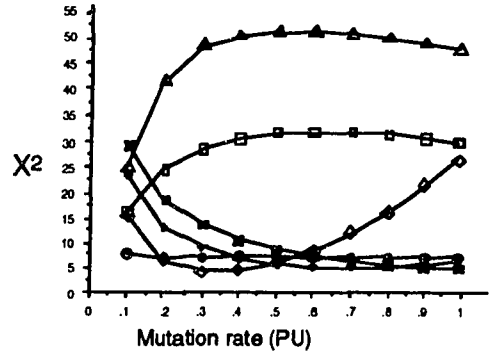


Figure 20a. Cage A1

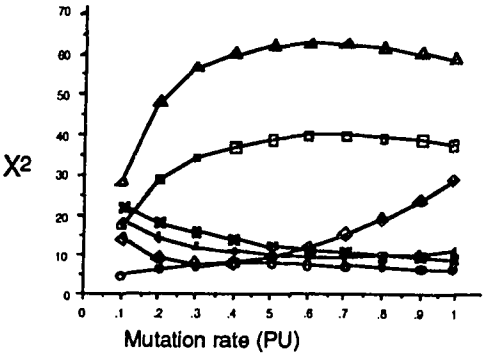


Figure 20b. Cage A2

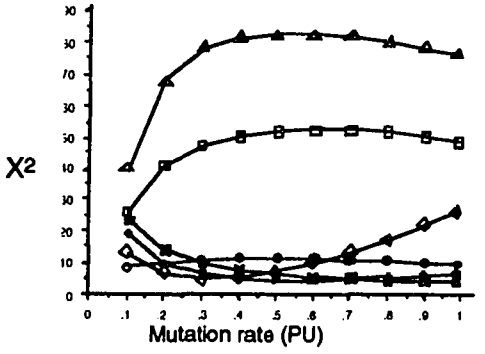


Figure 20c. Cage A3

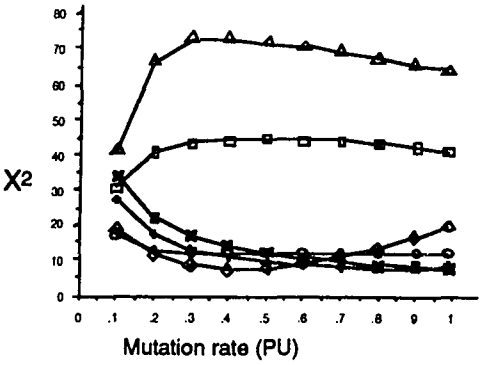


Figure 20d. Cage D1

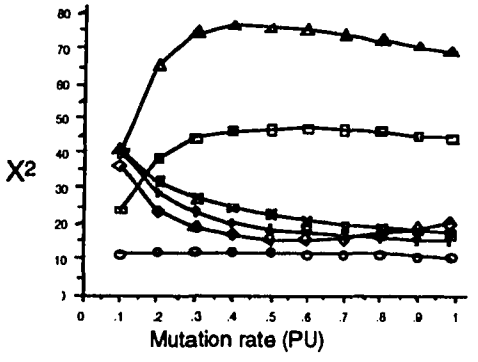


Figure 20e. Cage D2

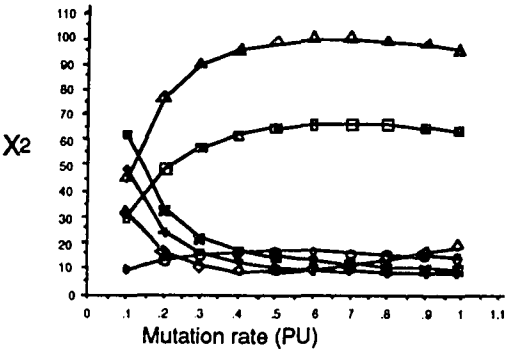


Figure 20f. Cage D3

Figure 20 a-f. Chi square values for fit of observed frequencies from all cages to the USCE Truncation model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

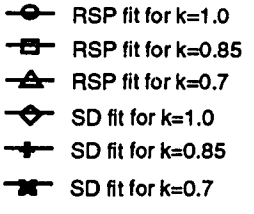


Table 4a. Minimum χ^2 for the fit of the observed values from cages A and D to the USCE Bounceback Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.58	4.46	0.2	7.74
	0.85	0.99	7.51	0.1	14.26
	0.7	0.99	13.08	0.1	19.98
A2	1	0.54	8.19	0.1	5.36
	0.85	0.99	11.62	0.1	15.13
	0.7	0.99	16.85	0.1	22.6
A3	1	0.54	4.58	0.1	9.77
	0.85	0.99	5.03	0.1	22.16
	0.7	0.99	7.42	0.1	31.76
D1	1	0.62	7.69	0.3	13.46
	0.85	0.99	10.19	0.1	27.69
	0.7	0.99	13.69	0.1	33.95
D2	1	0.99	15.31	0.1	11.69
	0.85	0.99	20.83	0.1	20.73
	0.7	0.99	26.05	0.1	32.27
D3	1	0.85	8.72	0.1	9.74
	0.85	0.99	12.18	0.1	25.64
	0.7	0.99	16.96	0.1	35.59

Table 4b. Joint best fit χ^2 for Cages A and D to the USCE Bounceback Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	1	0.57	4.83	20.65
A2	1	0.38	8.84	9.45
A3	1	0.35	5.54	12.79
D1	1	0.62	7.69	14.56
D2	1	0.99	15.31	14.79
D3	1	0.57	9.83	20.65

Table 4c. Joint best fit χ^2 for Cages A and D to the USCE Bounceback Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.99	7.51	30.4
A2	0.99	11.62	38.34
A3	0.1	26.18	22.16
D1	0.99	10.19	42.02
D2	0.99	20.83	44.4
D3	0.4	21.98	53.02

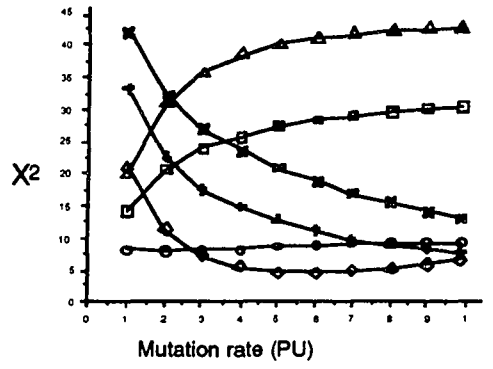


Figure 21a. Cage A1

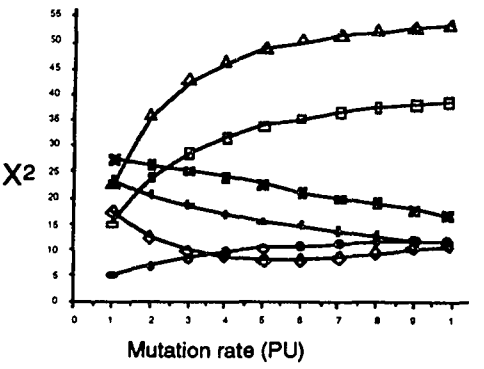


Figure 21b. Cage A2

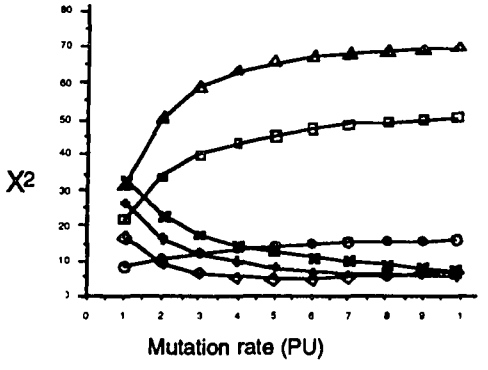


Figure 21c. Cage A3

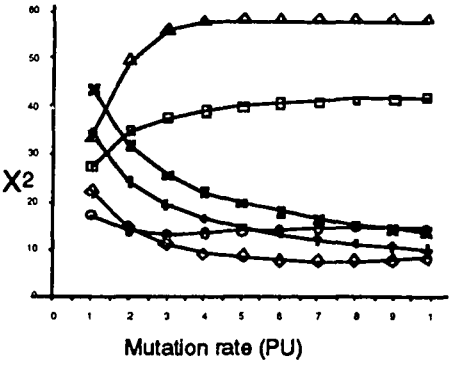


Figure 21d. Cage D1

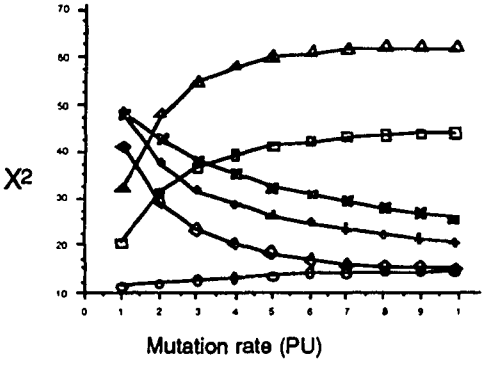


Figure 21e. Cage D2

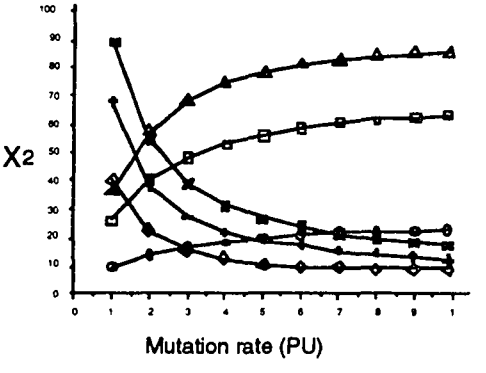


Figure 21f. Cage D3

Figure 21 a-f. Chi square values for fit of observed frequencies from all cages to the USCE Bounceback model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

RSP fit for k=1.0
RSP fit for k=0.85
RSP fit for k=0.7
SD fit for k=1.0
SD fit for k=0.85
SD fit for k=0.7

Table 5a. Minimum χ^2 for the fit of the observed values from cages A and D to the DSBR Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.1	4.47	0.41	7.53
	0.85	0.12	8.26	0.69	10.07
	0.7	0.13	11.39	0.77	9.98
A2	1	0.1	8.23	0.6	5.36
	0.85	0.1	13.95	0.85	7.01
	0.7	0.13	17.72	0.9	6.97
A3	1	0.1	4.53	0.58	8.67
	0.85	0.1	4.81	0.84	11.27
	0.7	0.13	5.23	0.89	11.19
D1	1	0.13	7.7	0.42	12.3
	0.85	0.16	11.42	0.7	17.09
	0.7	0.17	13.94	0.77	17.07
D2	1	0.17	15.3	0.5	11.38
	0.85	0.2	15.3	0.9	10.89
	0.7	0.2	15.7	0.91	10.96
D3	1	0.1	10.04	0.93	1.7
	0.85	0.1	11.53	0.99	6.63
	0.7	0.1	13.28	0.99	1.14

Table 5b. Joint best fit χ^2 for Cages A and D to the DSBR Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	1	0.1	4.47	8.96
A2	1	0.1	8.23	10.87
A3	1	0.1	4.53	14.74
D1	1	0.12	7.79	14.21
D2	1	0.17	15.3	14.6
D3	1	0.1	10.09	20.32

Table 5c. Joint best fit χ^2 for Cages A and D to the DSBR Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.1	8.54	38.16
A2	0.1	13.95	44.87
A3	0.1	4.81	61.39
D1	0.32	20.56	40.68
D2	0.37	24.76	35.85
D3	0.34	25.98	53.74

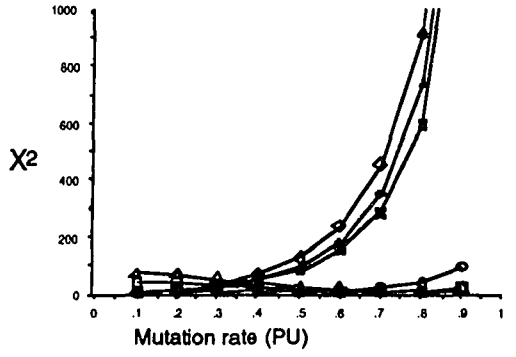


Figure 22a. Cage A1

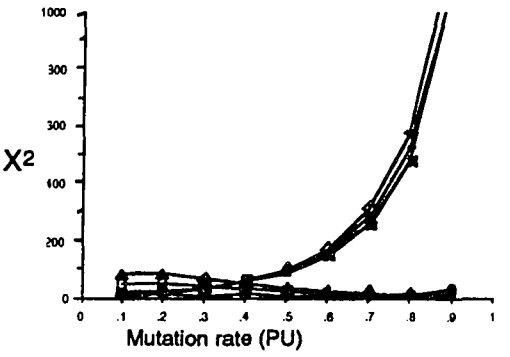


Figure 22b. Cage A2

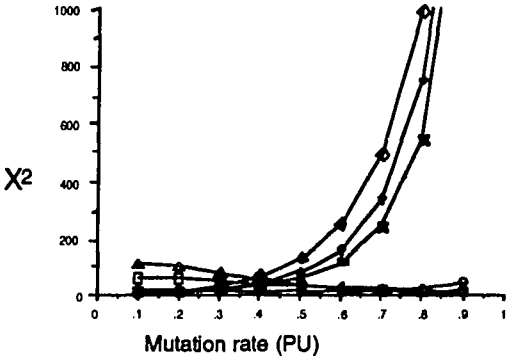


Figure 22c. Cage A3

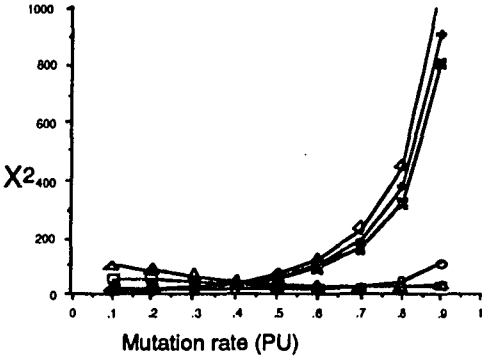


Figure 22d. Cage D1

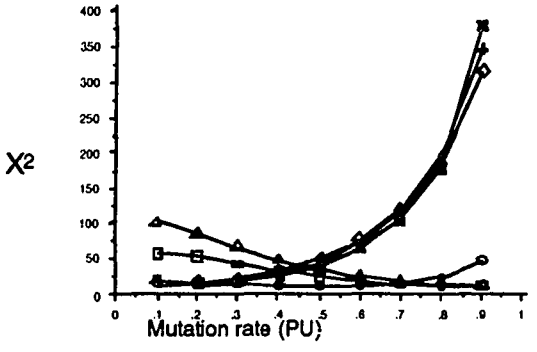


Figure 22e. Cage D2

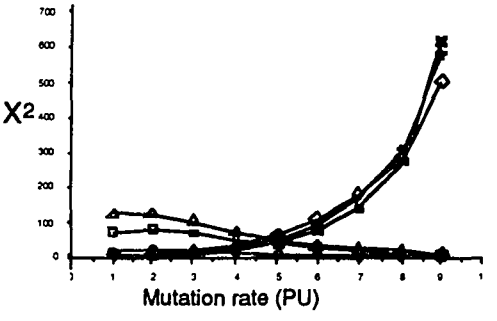


Figure 22f. Cage D3

Figure 22 a-f. Chi square values for fit of observed frequencies from all cages to the double strand Break Repair model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

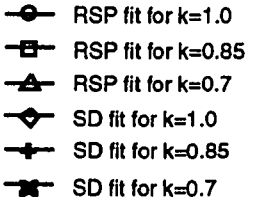


Table 6a. Minimum χ^2 for the fit of the observed values from cages A and D to the Hits Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.33	6.7	0.99	2.12
	0.85	0.37	11.22	0.1	10.32
	0.7	0.43	14.34	0.1	14.52
A2	1	0.29	10.92	0.1	3.9
	0.85	0.32	15.89	0.1	10.43
	0.7	0.37	19.4	0.1	16.63
A3	1	0.35	4.36	0.1	7.27
	0.85	0.41	5.25	0.1	14.4
	0.7	0.48	6.1	0.99	5.34
D1	1	0.36	11.54	0.99	2.79
	0.85	0.46	15.2	0.99	13.15
	0.7	0.56	16.8	0.99	14.97
D2	1	0.42	15.29	0.99	6.32
	0.85	0.44	15.59	0.99	10.1
	0.7	0.46	16.26	0.99	17.44
D3	1	0.4	8.51	0.1	6.96
	0.85	0.44	9.66	0.1	18.65
	0.7	0.5	10.53	0.1	27.36

Table 6b. Joint best fit χ^2 for Cages A and D to the Hits Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	1	0.31	6.76	8.88
A2	1	0.21	12.16	7.48
A3	1	0.28	5.24	13.38
D1	1	0.4	11.78	15.9
D2	1	0.42	15.29	14.55
D3	1	0.33	9.68	20.42

Table 6c. Joint best fit χ^2 for Cages A and D to the Hits Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.19	21.14	19.59
A2	0.1	22.27	10.43
A3	0.59	14.16	47.77
D1	0.18	33	26.45
D2	0.65	24.44	34.01
D3	0.59	15.41	64.96

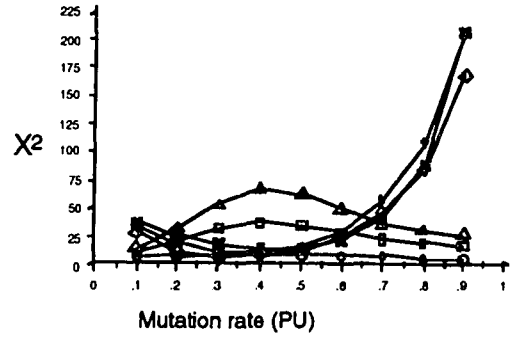


Figure 23a. Cage A1

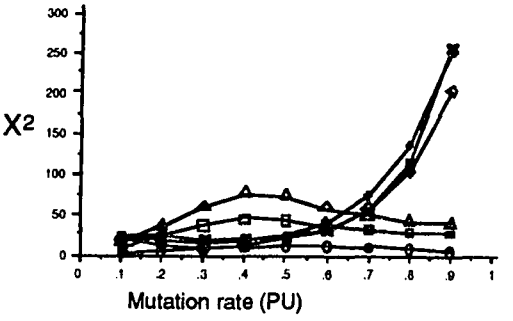


Figure 23b. Cage A2

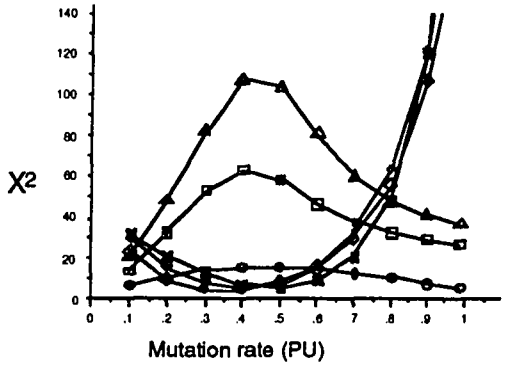


Figure 23c. Cage A3

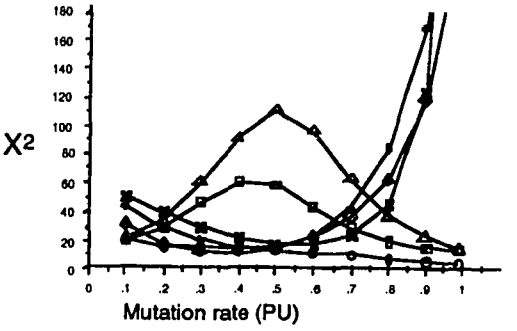


Figure 23d. Cage D1

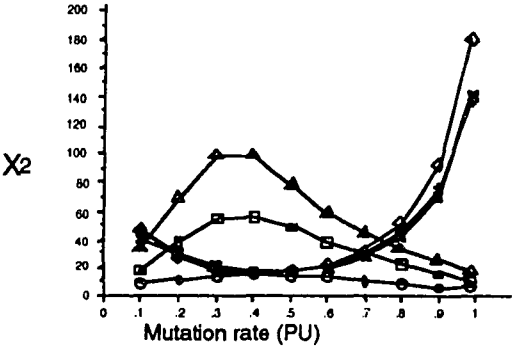


Figure 23e. Cage D2

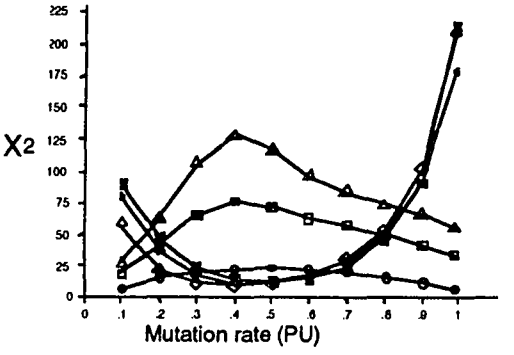


Figure 23f. Cage D3

Figure 23 a-f. Chi square values for fit of observed frequencies from all cages to the Hits model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

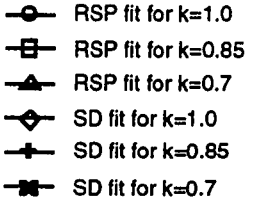


Table 7a. Minimum χ^2 for the fit of the observed values from cages A and D to the RSRsrm Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.1	88.83	0.1	12.78
	0.85	0.1	74.08	0.1	8.07
	0.7	0.1	61.26	0.1	4.88
A2	1	0.1	32.18	0.1	2.55
	0.85	0.1	29.26	0.1	1.52
	0.7	0.1	28.76	0.1	1.52
A3	1	0.1	64.4	0.1	6.82
	0.85	0.1	56.68	0.1	4.42
	0.7	0.1	48.88	0.1	3.21
D1	1	0.1	67.01	0.1	78.17
	0.85	0.1	64.73	0.1	60.91
	0.7	0.1	62.93	0.1	46.38
D2	1	0.1	93.6	0.1	10.89
	0.85	0.1	78.91	0.1	4.42
	0.7	0.1	62.8	0.17	2.78
D3	1	0.1	225.16	0.17	1.87
	0.85	0.1	195.83	0.31	1.69
	0.7	0.1	162.67	0.42	1.87

Table 7b. Joint best fit χ^2 for Cages A and D to the RSRsrm Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	0.7	0.1	61.26	4.88
A2	0.7	0.1	28.76	1.52
A3	0.7	0.1	48.88	3.21
D1	0.7	0.1	62.93	46.38
D2	0.7	0.1	62.8	2.78
D3	0.7	0.1	162.67	1.87

Table 7c. Joint best fit χ^2 for Cages A and D to the RSRsrm Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.1	74.08	8.07
A2	0.1	29.26	1.52
A3	0.1	56.68	4.42
D1	0.1	64.73	60.91
D2	0.1	78.91	4.42
D3	0.1	195.83	1.84

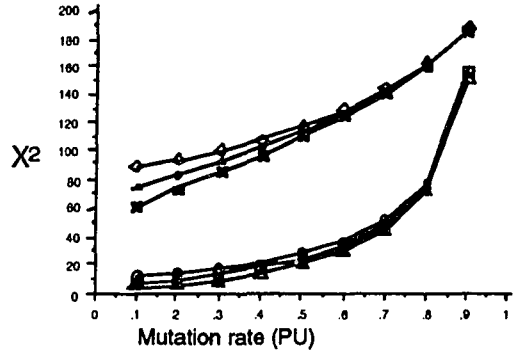


Figure 24a. Cage A1

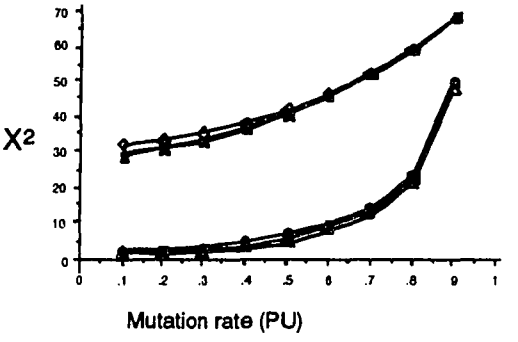


Figure 24b. Cage A2

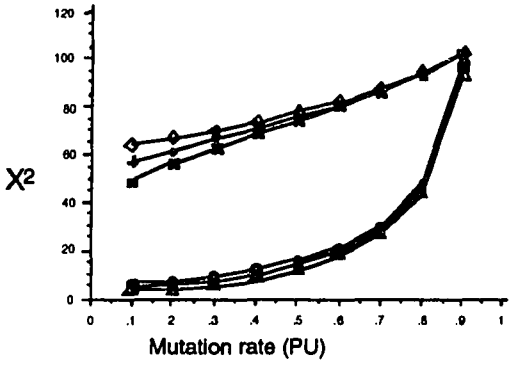


Figure 24c. Cage A3

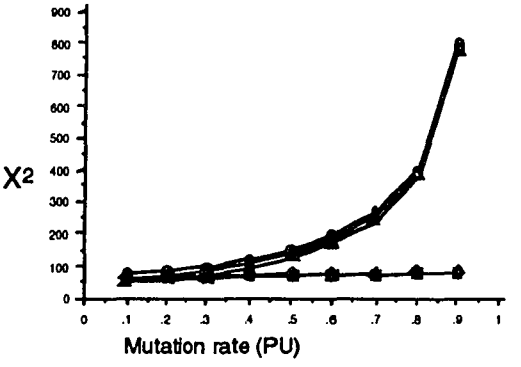


Figure 24d. Cage D1

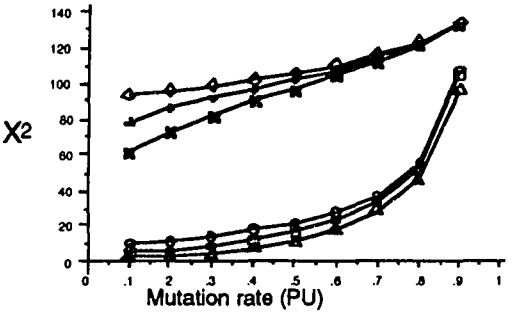


Figure 24e. Cage D2

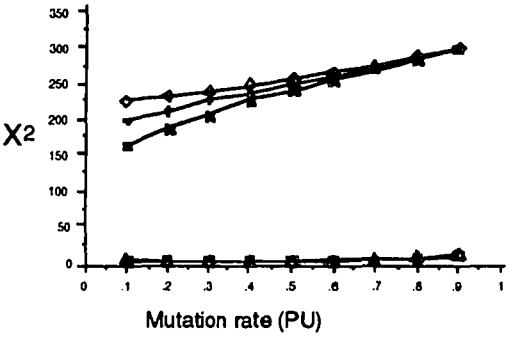


Figure 24f. Cage D3

Figure 24 a-f. Chi square values for fit of observed frequencies from all cages to the RSRsrm model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

RSP fit for k=1.0
RSP fit for k=0.85
RSP fit for k=0.7
SD fit for k=1.0
SD fit for k=0.85
SD fit for k=0.7

Table 8a. Minimum χ^2 for the fit of the observed values from cages A and D to the RSRamm Model where k= 1, 0.85, and 0.7. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2.

		Sd		*Rsp*	
Cage	k	PU	X2	PU	X2
A1	1	0.1	9.42	0.22	7.75
	0.85	0.1	10.17	0.63	8.04
	0.7	0.1	12.56	0.7	6.94
A2	1	0.1	13.01	0.34	5.86
	0.85	0.1	16.39	0.77	4.34
	0.7	0.1	19.71	0.84	3.73
A3	1	0.1	8.16	0.32	9.37
	0.85	0.1	9.95	0.76	6.76
	0.7	0.1	6.57	0.7	6.31
D1	1	0.1	9.28	0.22	12.55
	0.85	0.1	11.87	0.64	14.14
	0.7	0.1	14.31	0.7	12.28
D2	1	0.1	15.46	0.28	11.73
	0.85	0.1	15.34	0.69	8.83
	0.7	0.1	15.8	0.77	8.24
D3	1	0.1	9.55	0.99	2.06
	0.85	0.1	9.03	0.97	1.28
	0.7	0.1	9.78	0.77	1.33

Table 8b. Joint best fit χ^2 for Cages A and D to the RSRamm Model. K and PU are the combination that gives the smallest χ^2 for both *Sd* genotype frequencies and *Rsp* genotype frequencies.

Cage	k	PU	SD X2	*RSP* X2
A1	1	0.1	9.42	9.48
A2	1	0.1	13.01	12.35
A3	1	0.1	8.16	16.54
D1	1	0.1	9.28	14.68
D2	1	0.1	15.46	14.54
D3	1	0.1	9.55	23.42

Table 8c. Joint best fit χ^2 for Cages A and D to the RSRamm Model where k= 0.85. For *Sd* genotype frequencies and *Rsp* genotype frequencies, PU is the value that gives the smallest χ^2

Cage	PU	SD X2	*RSP* X2
A1	0.1	10.17	40.84
A2	0.1	16.39	51.08
A3	0.1	9.95	66.64
D1	0.18	21.41	41.53
D2	0.21	25.02	36.89
D3	0.22	25.83	56.15

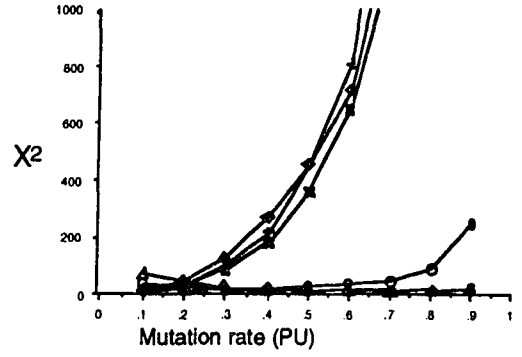


Figure 25a. Cage A1

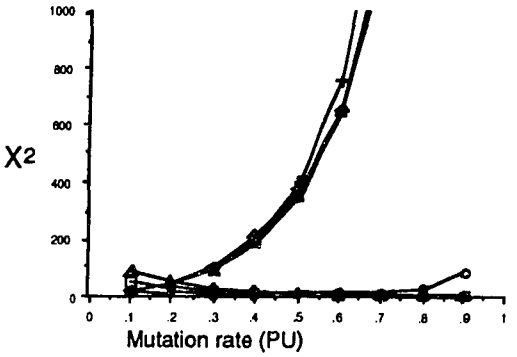


Figure 25b. Cage A2

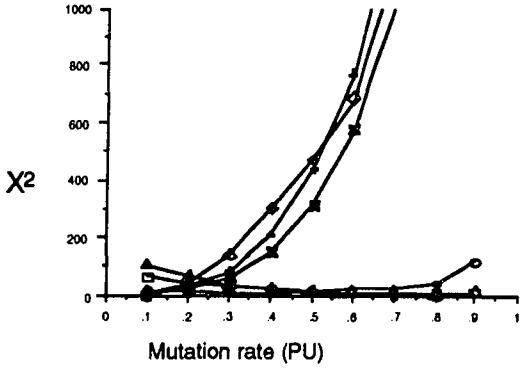


Figure 25c. Cage A3

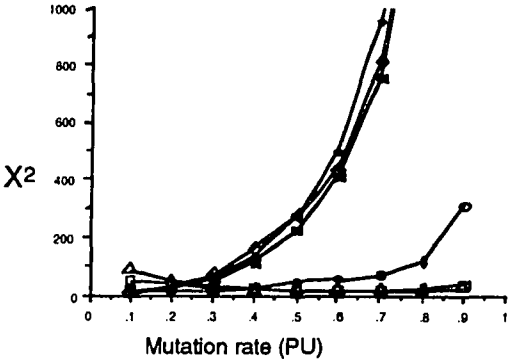


Figure 25d. Cage D1

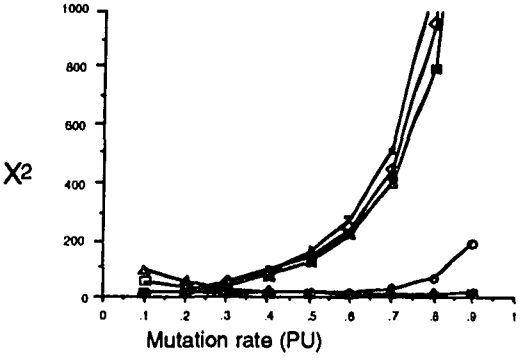


Figure 25e. Cage D2

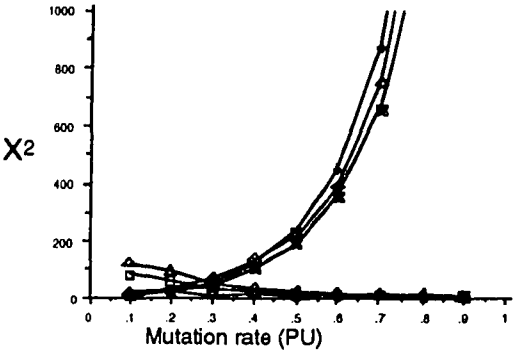


Figure 25f. Cage D3

Figure 25 a-f. Chi square values for fit of observed frequencies from all cages to the RSRamm model for 9 values of mutation rate(PU). X^2 is the fit of observed genotypic frequencies to theoretical values for k=1, k=0.85 and k=0.7.

RSP fit for k=1.0
RSP fit for k=0.85
RSP fit for k=0.7
SD fit for k=1.0
SD fit for k=0.85
SD fit for k=0.7

patterns. In general, the *Sd* fit is best to intermediate values of PU while the *Rsp* fits are best at low values of PU for all k. The A cages give a best joint fit for k= 1.0 and PU of about 0.3 for the truncation model and slightly higher PU for the bounceback model. The D cages also give best joint fit at k=1 and PU ranging from 0.42 to 0.6 for truncation and 0.57 to 0.99 for bounceback. The test is for goodness of fit of 2 non-independent data sets with a total of 6 trials with 5 classes each to an expected distribution. Degrees of freedom are difficult to assess for this analysis, so statistical significance is imprecise. The joint χ^2 values are all larger than 12.5. While this is much lower than seen in some of the other models, it is apparent that the null hypotheses should be rejected.

3.6.2 χ^2 Analysis of the DSBR Model

The double strand break repair model gives the best fits are at low values of PU. χ^2 values were very high for SD fits for PU > 0.7. The *Rsp* fits are best in the intermediate PU's (Figures 22a-f and Table 5). The best joint χ^2 values are all greater than 13.5 and so the null hypothesis is rejected.

3.6.3 χ^2 Analysis of the Hits Model

The *Sd* moderated model gives quite different results than the previous models (Figures 23a-f and Table 6). The χ^2 values for *Rsp*

frequencies are highest for intermediate values of PU and lowest on either end of the range, with the lower end being slightly higher for k=1.0 and lower for k= 0.85 and k=0.7. The χ^2 values for *Sd* are lowest at about PU = 0.35 for all values of k. The best joint χ^2 values are all greater than 15.5 and so the null hypothesis is rejected.

3.6.4 χ^2 Analysis of the RSR Model

The two different rules on misalignment size produce very different results. The single repeat misalignment rule (srm) model gives χ^2 values that are extremely high for all PU's and k's, with χ^2 for both *Sd* and *Rsp* being lowest with lower PU values (Figures. 24a-f and Table 7). The lowest best joint χ^2 is 30.28 and the rest are all greater than 50. The null hypothesis is rejected for this model.

The allele maximum misalignment rule (amm) produces χ^2 values for *Sd* that are lowest at low PU values (Figures 25a-f and Table 8). The χ^2 values for *Rsp* are lowest at intermediate values of PU. The best joint χ^2 values are all greater than 15 and so the null hypothesis is rejected.

The values that gave the best fit for each model were used to generate graphs of single simulation runs for 20 generations. The χ^2 values in Table 9 are means of the best joint fit χ^2 for the three A cages. The mean allele frequencies of the A cages were superimposed on each graph.

Table 9. Values of drive strength (k) and mutation rate (PU) that produced the minimum χ^2 for each simulation model. Figure number refers to the associated graph for that model.

Model	k	PU	χ^2	Figure number
USCE truncation	1.0	0.30	14.01	26
USCE bounceback	1.0	0.38	22.37	27
DS Break Repair	1.0	0.10	17.27	28
Hits	1.0	0.25	17.97	29
RSRsrm	0.7	0.10	49.50	30
RSRamm	1.0	0.10	23.58	31

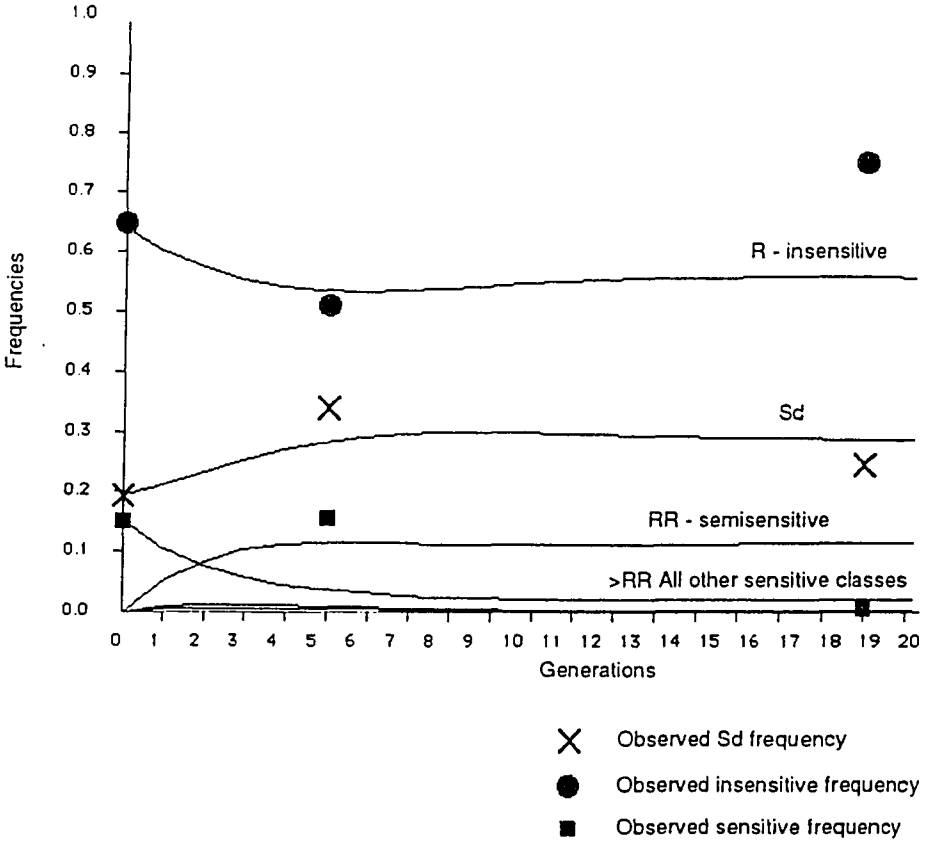


Figure 26 USCE Truncation model with k = 1.0 and PU = 0.30

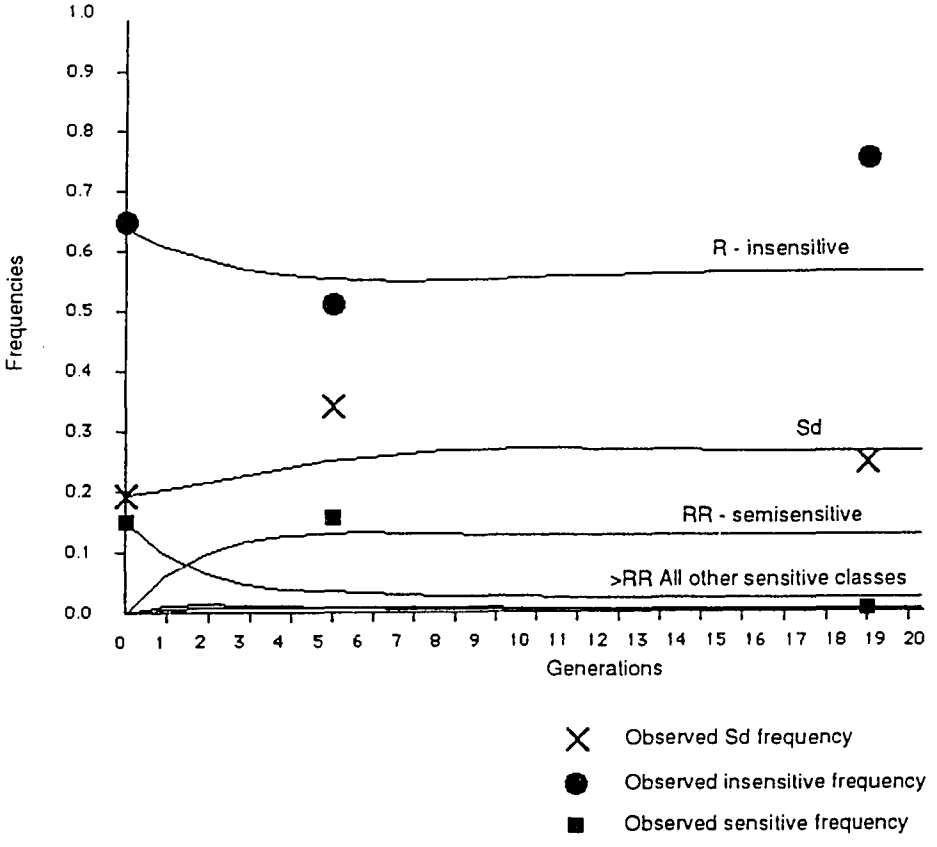


Figure 27 USCE Bounceback model with k = 1.0 and PU = 0.38

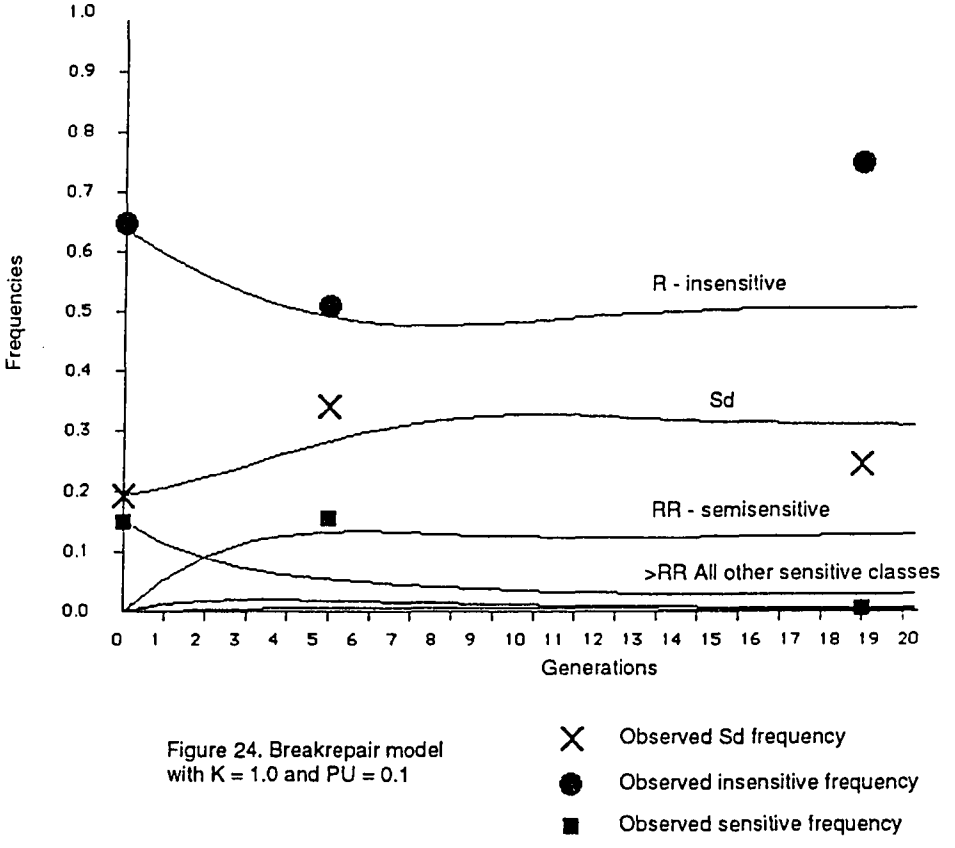


Figure 24. Breakrepair model with K = 1.0 and PU = 0.1

Figure 28 Double Strand Break Repair model with k = 1.0 and PU = 0.10

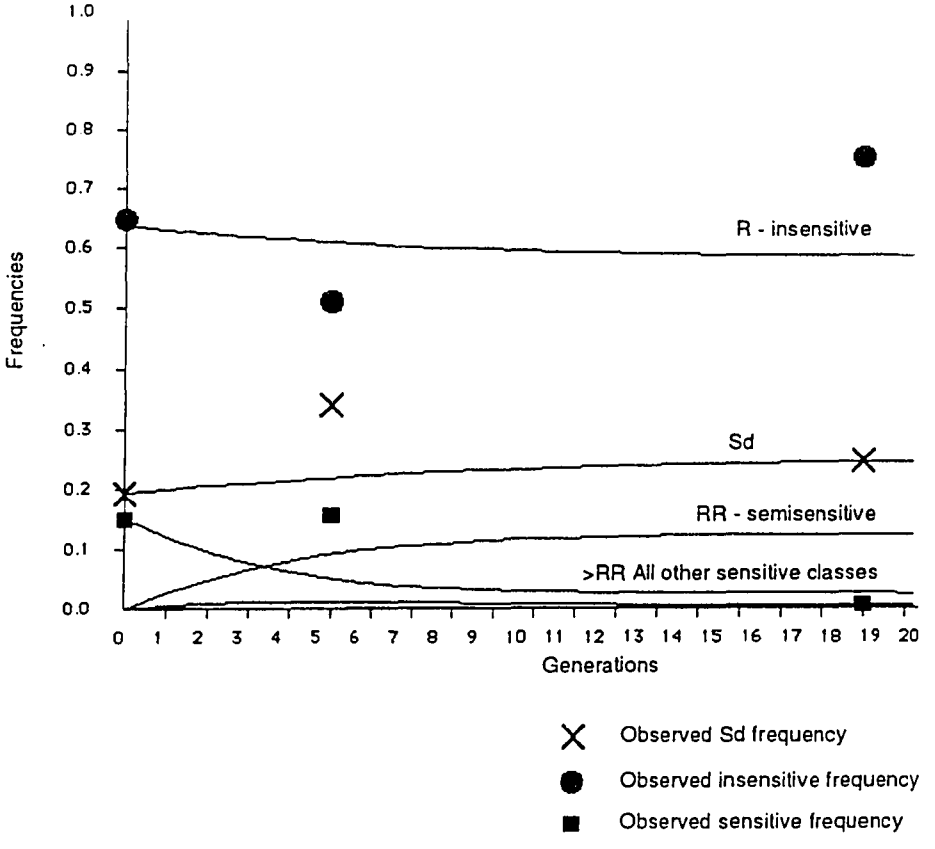


Figure 29 Hits model with k = 1.0 and PU = 0.25

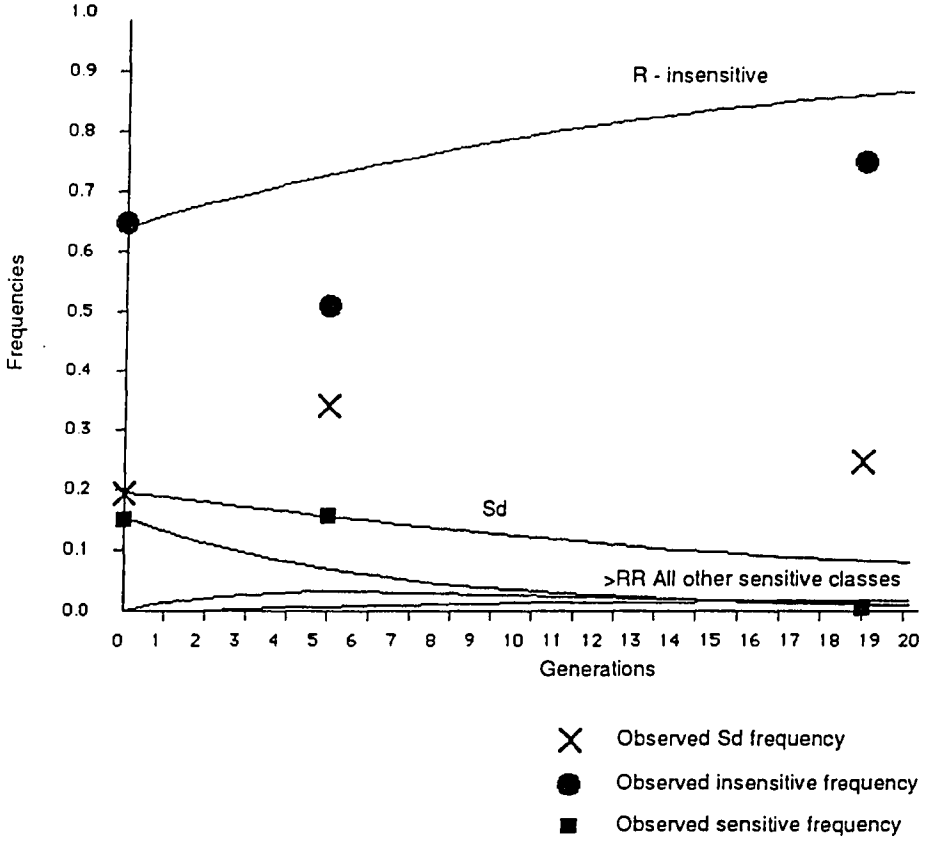


Figure 30 RSRsrm model with k = 0.7 and PU = 0.10

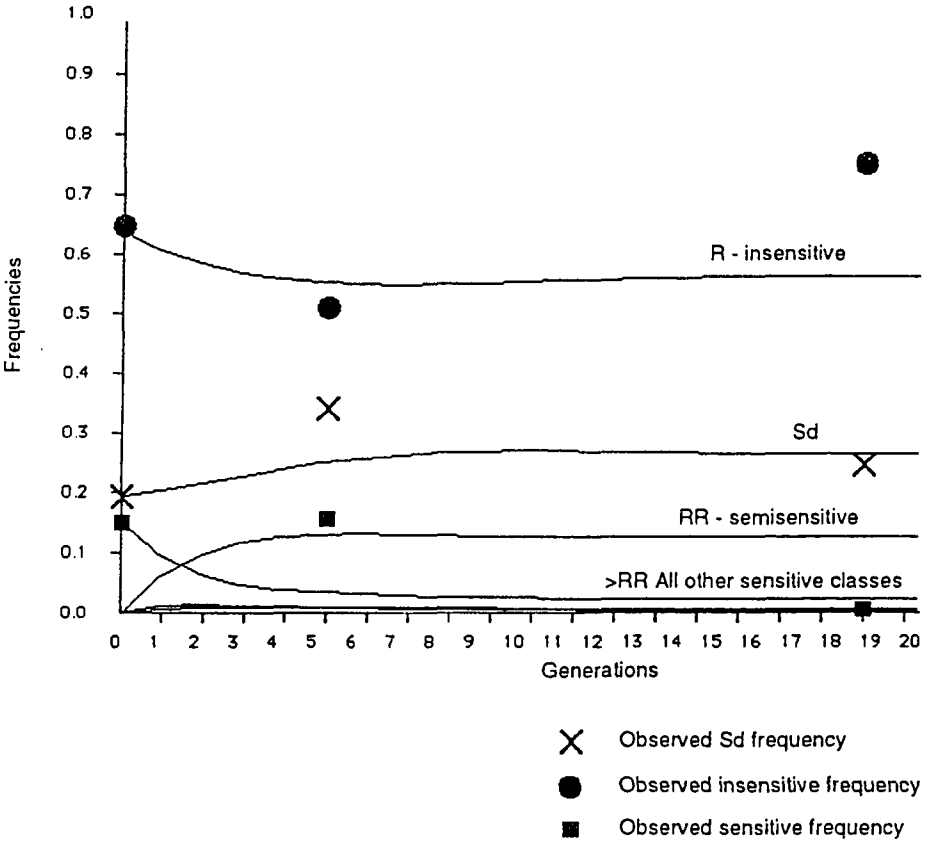


Figure 31 RSRamm model with k = 1.0 and PU = 0.10

As discussed earlier, statistical interpretation of these results is problematical. Ignoring the non-independence of the data would give 5 data classes (three *Sd* phenotypes and two *Rsp* phenotypes), for a maximum of 3 degrees of freedom. For 3 degrees of freedom, the ritical value of the χ^2 distribution for a=0.95 is 7.81. Therefore, even the most liberal interpretation of the χ^2 statistic would reject the hypotheses that the observed values fit the expected distribution for any of the six models. Confounding factors would reduce the degrees of freedom and a more conservative interpretation would reject even more strongly.

In some instances, the fit of the observed data to the predicted seems close. However, two more caveats are required. First, the *Sd* used in the cages has a measured drive strength against the sensitive Responder used of about k=0.94. For all models and PU values, the fit of the observed values to those predicted with intermediate k (i.e., less than 1.0 or greater than 0.7) is worse than the best values, which were always at either k=1.0 or k=0.7. Thus the simulations are worse predictors of the observations given the parameters actually used in the cages.

Secondly, the parameterizations that fit best are those that lead to loss of the sensitive Responder haplotype. While this is what happened in the cages, it not what is observed in natural populations. The hypothesis that these processes are producing the dynamics seen in nature is therefore not supported by these results.

3.7 Experimental Cage Results for Cage F: *Rsp* Frequencies

Among the three F cages with no *Sd*, a total of 12 samples were assessed for *Rsp* frequencies using Southern blot analysis. Results are summarized in Table 10.

Table 10: Responder frequencies are in the form: Freq± Var (sample size). The variance is the variance of the estimate of frequency from the MLE. Sample sizes are number of scorable lanes for the *Rsp* Southern blots.

Cage	allele	11/18/94	2/3/95	4/11/95	9/7/95
F1	sens	0.1056	0.338	0.351	0.312
	ins	0.8944	0.662	0.649	0.688
		±0.0025 (20)	±0.0069 (19)	±0.0069 (19)	±0.0069 (19)
F2	sens	0.3175	0.423	0.334	0.40
	ins	0.6825	0.577	0.666	0.60
		±0.0069 (19)	±0.0069 (19)	±0.0069 (19)	.0069 (19)
F3	sens	0.195	0.423	0.330	0.39
	ins	0.805	0.577	0.670	0.61
		±0.0069 (19)	±0.0069 (19)	±0.0069 (19)	±0.0069 (19)

3.8 Analysis of F cage Results

Since this set of cages contains no Segregation Distorter alleles, the only force that could produce a directional change in allele

frequency is a fitness differential between the sensitive and insensitive Responder bearing chromosomes. The observation was made (Table 10) that for all three F cages, *Rsp*^{i} was lower in frequency at the last collection than the first. This is consistent with selection favoring the sensitive allele.

The level of selection was assessed using the method of minimum χ^2, after Wu, *et al* (1988). They were measuring a fitness difference between a standard lab *Rsp*^{s} and a derivative deleted for the Responder region. They compared observed frequency changes in cage populations with theoretical predictions based on various values of selection. The the value of selection that gave the best fit was considered the best estimate of selection in the cages. Using the same approach here makes the results directly comparable.

A series of iterations were performed to generate expected frequencies at generation times corresponding to those of the actual collections. Theoretical values were obtained using the formula

$$\Delta q=-spq[q+h(p-q)]/(1-2pqhs-q^2s)$$

where p and q are allele frequencies, h is the degree of dominance and s is the fitness differential.

Wu *et al* (1988) found that the best fit was obtained assuming complete dominance (h=0). Substituting this value, the formula becomes

$$\Delta q=-spq^2/(1-q^2s).$$

χ^2 values were calculated for each tested value of s by comparing the expected genotype frequencies with the observed genotype frequencies. The value of s that produced the minimum χ^2 for each cage was taken as the best estimate of s.

The results are summarized in Table 11 and presented graphically in Figures 32-34.

Table 11. Estimates of s and χ^2 values for three F cages.

Cage	s	χ^2
F1	0.087	3.17258
F2	0.025	0.507
F3	0.068	1.768

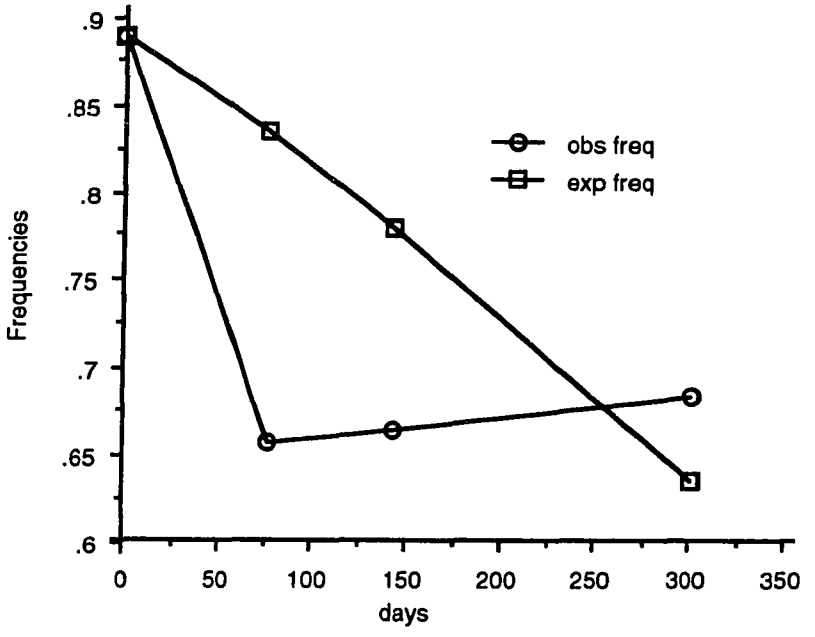


Figure 32. Graph of theoretical frequencies for s = 0.087 and observed frequencies of insensitive responder cage F1.

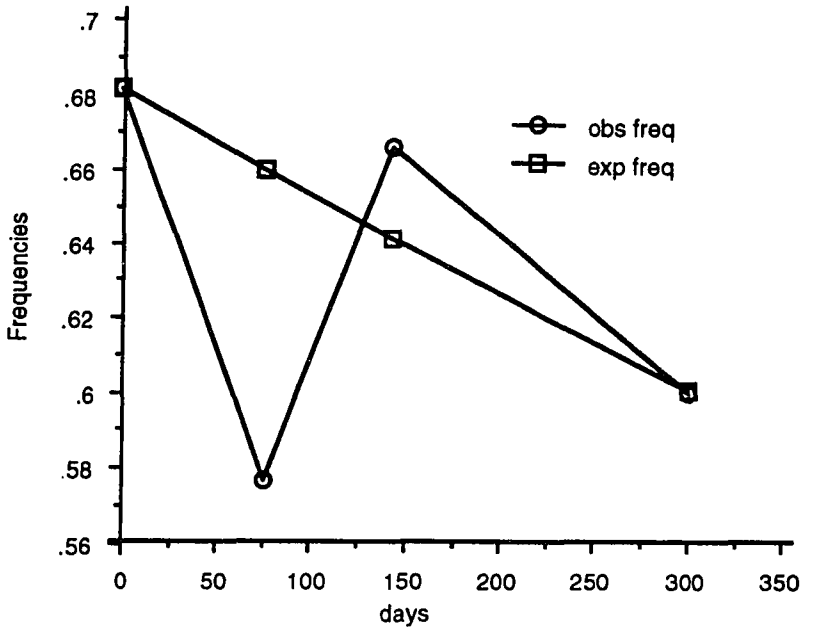


Figure 33. Graph of theoretical frequencies for s = 0.025 and observed frequencies of insensitive responder cage F2.

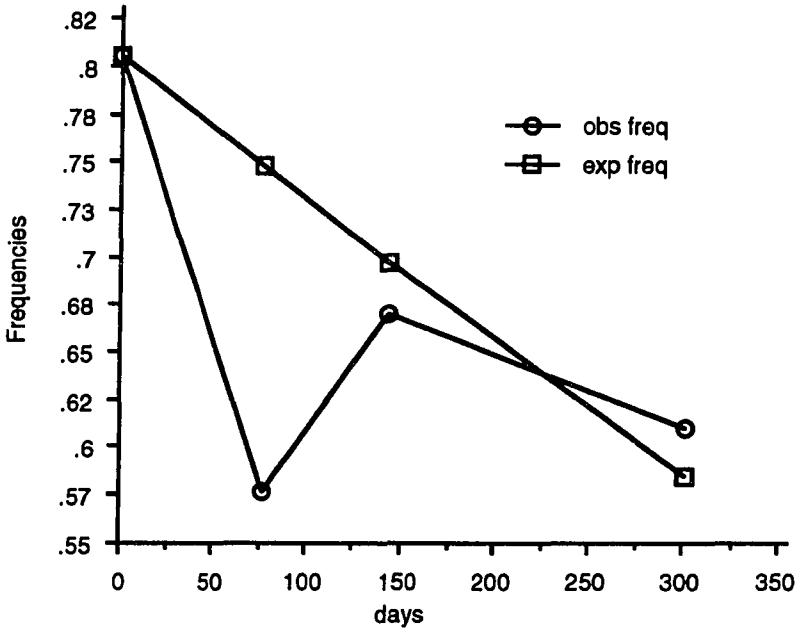


Figure 34. Graph of theoretical frequencies for s = 0.068 and observed frequencies of insensitive responder cage F3

Chapter 4
Discussion

4.1 Mutation - Drive Balance Produces Stable Polymorphism in Simulations

The simulations demonstrate that mutational processes can produce a stable polymorphism. This is an important result because it opens the possibility that mutation may be a significant factor in the polymorphism seen in nature. As discussed in the introduction, repetitive DNA has properties that make expansions and contractions common and there are several well known instances where the dynamic nature of these sequences have a profound effect on phenotypes. Furthermore, one likely role of repetitive DNA in the cell is in chromosome structure. This means that the possibility exists that such sequences may be candidates for being targets of meiotic drive. Together, these observations suggest that a system such as Segregation Distorter may not be a unique case and that a balance between meiotic drive and mutation of a repetitive target could maintain polymorphism.

The simulations show that the rates of change required to maintain a polymorphism are quite high. This may be plausible, given that expansions and contractions of repetitive DNA are due to special mechanisms that may indeed have high rates. Indeed,

mutation rates for many of the human diseases associated with repetitive DNA are known to be high.

A second major result of the study is that the polymorphisms predicted by the simulations are not those seen for the Segregation Distorter system in nature. In particular, for most reasonable parameterizations, the equilibrium frequencies of *Rsp*^{s} are much lower than observed in nature. This does not negate the importance on mutational process at the Responder locus, rather, it simply implies that mutation alone may not be enough to explain the natural polymorphism. The computer simulations were designed to simulate the system stripped of all other variables and so to test for the effect of mutation alone. Apparently other factors are important in the natural population that were intentionally excluded from the computer models.

4.1.1 Responder Alleles Appear to be Stable

If *Rsp*^{s} were being produced de-novo from *Rsp*^{i} at an appreciable rate, chromosomes with higher copy number should arise in populations of only *Rsp*^{i}. In the forensic use of minisatellite DNA fingerprints, the range of sizes that is considered useful is from about 25Kb down to 4Kb (Jefferys et al., 1989) and the mean number of bands for many commonly used probes is 36. Jefferys et. al. (1989) found that band shifts due to a change in repeat copy number arise at an estimated rate of 0.006 per offspring band. At this rate, a

new band is detected in about 25% of the children of a particular parent. The Responder fingerprints discussed here are similar in many respects to the minisatellite fingerprints. A large portion of a Responder fingerprint falls into the same size range. We should have been able to detect a change of a single repeat unit (240bp) given our level of resolution. If the Responder copy numbers were changing at a rate comparable to that seen by Jefferys, we should have seen it given the large number of alleles identical by descent that were examined.

It is important to emphasize that the experiments presented here were not designed to detect small changes in copy number. The changes that we are interested in for this study are those that involve enough copies to cause a change in allelic state. This degree of change would require at least several tens of copies. Forty new copies would be about 10Kb of new material. This would very likely have been detectable on our standard Southern blot. The isoline bottle stocks of Rsp^i that were maintained for this study were isogenized for Rsp^i in 1992 and maintained at large population sizes since then. If Rsp^i were becoming Rsp^s at an appreciable rate, new alleles should have been seen. However, in recent checks of the bottle stocks using Southern blot analysis (data not shown), there were only Rsp^i with the same pattern that had been observed from the beginning of the study.

This observation is not new, as Rsp^i bearing stocks have been maintained in the lab for many decades. An alternate hypothesis is

that the action of *Sd* causes an increase in sensitivity. Lab stocks are usually been maintained in the absence of *Sd*. In the present study, one of the cages, D2, lost Rsp^s early in the study (before February 1995) yet continued to segregate *Sd* and Rsp^i. If *Sd* causes a high rate of change in the *Rsp* repeat sequence, it should have been seen in this cage, given the large population size of the cage and the relatively long time of the study. However, after several generations, a sample of 30 flies showed no new alleles. This is by no means an exhaustive examination of the cage, and new Rsp^s alleles could be present at low frequencies.

Based on binomial sampling, a sample size of 30 gives 95% confidence of detecting a rare allele present at any frequency down to 0.0487. While we can not rule out completely the production of new Rsp^s alleles, we can say that Rsp^s is apparently not being produced de-novo at a high rate. In particular, it is clear that allele change is not happening at a rate high enough to be significant in balancing the effect of meiotic drive.

4.1.2 Segregation Distortion Dominates Cage Dynamics

The cage frequency changes suggest that meiotic drive is the dominant force acting on allele frequencies in the cages. With no mutation and only meiotic drive acting, single runs of the simulations show that, in the early generations, Rsp^s decreases in frequency continuously. *Sd* briefly increases in frequency while Rsp^s is common

but then decreases in frequency as *Rsp*^{s} becomes rare. When *Rsp*^{s} is lost, *Sd* has no further advantage in the population and any further change may be attributed to the reduced fertility of SD/SD males. This is compatible with what was observed for all six cages.

The cage experiments were designed to take a first look at whether there is evidence of the rates of mutational activity that would be required to control *Responder* allele frequencies. Our results demonstrate that mutation is apparently not the sole factor in determining *Responder* allele frequencies. We cannot rule out the possibility that mutation in repeat copy number does occur at a small rate. Therefore, mutation may contribute to the maintenance of the polymorphism. It is quite plausible that a combination of factors, especially fitness differences, operate together to keep the SD components present in populations. This study did not address that question, but rather was intended solely to test whether mutation alone was a sufficient explanation.

4.2 Fitness Advantage of *Rsp*^{s} over *Rsp*^{i}

The analysis of allele frequency in the F cages revealed a fitness advantage of Kona 3 *Rsp*^{s} over Kona 13 *Rsp*^{i}. This is similar in directionality to the results obtained by Wu et al (1988) but is quantitatively much lower. This supports the hypothesis that Responder provides a small fitness advantage but suggests that it is less important than suggested in the previous study.

We tested only two native chromosomes. They were drawn from the same population and maintained with as much background variability as possible, being selected for *Responder* allele only. The sublines represent two random draws from the population. The fact that the fitness difference in this single comparison is in concordance with previous studies is highly suggestive but is by no mean a conclusive result. Although the two isolines were randomized for background genetic variability, it is an inescapable fact that isogenizing lines in this manner creates a linkage disequilibrium that persists for some time after the lines are mixed. Therefore, while we measured a change in Responder allele frequencies that was attributed to a fitness difference, it is quite possible that the actual fitness difference is due to loci linked to Responder. A more exhaustive demonstration would require the testing of several chromosomes drawn from nature.

In the present study, the fitness difference observed is too low to counter the effects of drive alone, as evidenced by the extinction of Rsp^s in all the test cages where SD was present. In terms of its role in maintaining the polymorphism of the SD system in nature, the level of selection measured is also too low to be the sole reason for the polymorphism. However, the possibility remains that a low but consistent fitness advantage for Rsp^s is one of several factors that contributes to its persistence.

4.3 The Population Cages Did Not Reach a Balanced Polymorphism.

The population cages did not replicate what is observed in nature with regard to *Responder* and SD frequencies. There are several possible reasons for this. It is clear that some significant factor present in nature was absent from the cages. The possible factors fall into two categories: environmental and genetic.

The first possibility is that the experiments failed to include some important factor of the environment. The cage experiments were conducted in a laboratory incubator where temperature, light, humidity and food are carefully controlled. Under these conditions, the populations did not reach a stable polymorphism, as Rsp^s was lost from all cages. It is entirely possible that the larger repeat array of Rsp^s provides some fitness advantage in the highly variable natural environment that is not manifested in the laboratory. While this is quite possible, it would be very difficult to design an experiment to assess the difference.

The second possibility is that some important genetic element was missing from the cages. The second chromosomes in the cages were all characterized as to viability, fertility, sensitivity and drive strength. The 1st, 3rd and 4th chromosomes were not isogenized and presumably supply a random genetic background similar to that found in nature. However, they do represent a bottlenecked population and genetic variability in these chromosomes is expected

to be lower than in nature. If natural suppressors of meiotic drive play a significant role in maintaining a balance, it would not be evident in these experiments since such suppressors were specifically excluded. This is discussed in detail in the subsequent section. The possibility also exists that there are other, unknown genetic factors that contribute to the polymorphism. Interaction with alleles at other loci that were absent from the experimental populations may be important. A similar set of experiments with much more genetic variability at all loci could at least begin to address this possibility.

4.4 The Possible Role of Modifiers of Drive.

Apparently, one or more components present, and critical, in nature were absent from the population cages. While there are several potential candidates, one set of genetic elements usually found in nature that were specifically excluded from the cages are the natural suppressors of drive.

Theoretical studies have shown that the evolution of genetic suppressors can limit the spread of driving alleles (e.g.. Stalker, 1961; Feldman and Otto, 1991; Bengtsson and Uyenoyama, 1990; Haig and Grafen, 1991; Hartl, 1975; Liberman, 1990). The fact that meiotic drive theoretically reduces the mean fitness of a population means that meiotic drive produces strong selection on the organism, at the diploid level, to develop mechanisms to suppress the effects of drive.

This was demonstrated in a series of studies (described in Lyttle, et. al., 1993) in which a pseudo Y-drive system was created. Sex chromosome distortion imposes a much stronger negative effect of meiotic drive than that in autosomal systems since extreme sex ratio distortion rapidly leads to population extinction. In this system, suppressors of drive would be expected to evolve rapidly. Lyttle (1979) found that most population cages went to extinction in 10-15 generations but a few persisted. In these, the drive strength gradually dropped from the initial value of k=0.94 to k=0.83, over about 40 generations. This was attributed to the effect of polygenic, quantitatively acting suppressors, each of small effect, distributed across both the X and autosomes (Lyttle, 1979).

The presence of suppressors of meiotic drive in natural populations has been known from the early days of the study of SD (Hiraizumi, Sandler and Crow, 1960). Hartl (1970) found both autosomal and X-linked suppressors in a population of *D. melanogaster* in Madison Wisconsin at frequencies of 45% and 85% respectively. Similarly high frequencies of suppressors have been found in other populations (Hartl and Hartung, 1975; Hiraizumi and Thomas, 1984; Kataoka, 1967).

Although Rsp^s chromosomes are found in typical populations at very high frequencies, the high frequencies of suppressors means that the net drive may be effectively much lower in natural populations than that suggested in laboratory tests of *SD* drive strength. For the cage experiments presented here, the lines were

chosen by their behavior in tests of meiotic drive. This had the effect of automatically screening out any lines that were carrying strong unlinked suppressors. Thus, while the SD used had only moderate drive strength, it was free from the influence of any strong suppressors of drive in the cages. Its drive strength therefore may have been effectively stronger than it is in nature, resulting in the extinction of *Rsp*^{s} in the cages.

I speculate that the evolution of suppressors of drive in natural populations may be a key feature that determines the polymorphism that is observed. It is plausible that a mutation-selection balance maintains *Rsp* alleles and that a shifting balance between SD advantage and suppression of SD maintains SD.

Including all three factors, selection, mutation and suppressor evolution makes a model that is complicated but Occam's Razor says to prefer the simplest hypothesis that explains the observations. So far, none of these three alone seems to be a sufficient explanation and so it becomes necessary to devise more complex explanations.

Appendix
Molecular Protocols

A.1 Plasmid Constructs: PGN156 and H0

The PGN156 plasmid used in this study was provided by Pat Powers of University of Wisconsin. Plasmid PGN156 contains a 5 Kb EcoR1 fragment in a pGeml vector (from Promega-Biotech). The 5kb EcoRl fragment includes the *Sd*+ region and is unique to *Sd* and hybridizes to both the *Sd* and *Sd*+ alleles from genomic DNA.

The H_0 plasmid used in this study was provided by C-I Wu. Plasmid H_0 is a 2.5 Kb EcoR1 fragment in a pUC9 vector (Wu, et al 1988).

A.2 Bacteria : Liquid Cultures

This protocol is taken from Current Protocols in Molecular Biology. Bacteria containing plasmids PGN156 and H_0 were grown by this method.

Materials

2x YT: 16 g Bacto-tryptone, 10 g Bacto-Yeast Extract, 5 g NaCl, in 1 L of ddH_2O. Autoclaved. Stored at 4° C. 20% glucose: Filter sterilized. Stored at 4°C .

Ampicillin 25 mg/ml: Filter sterilized. Stored at -20°C in 1 ml aliquots.

Methods

Each bacterial culture was grown in 40 ml of 2xYT media supplemented with 40 ul of 20% glucose and 80 ul ampicillin (25 mg/ml) in a 250 ml Erlenmeyer flask overnight on a shaker at 37° C .

A.3 Plasmid DNA Isolation

The protocol for isolating plasmid DNA is adapted from Vollmer and Yanofsky (1986). Plasmid DNA from PGN156 and H_0 were isolated by this method.

Materials

Lysozyme (10 mg/ml): Stored at -20° C in 1 ml aliquots.

95% ethanol: Stored at -20° C .

7.5 M NH_4OAc: Filter sterilized. Stored at room temp.

GTE: 50 mM glucose, 25 mM Tris-HCl pH 8, 10 mM Na EDTA. Autoclaved. Stored at room temp.

0.2N NaOH in 1% SDS: 80 ul 5N NaOH, 1720 ul ddH_2O, 200 ul 10% SDS. Made fresh just prior to use.

5 M KOAc: 90 ul 5M KOAc, 42.3 ul ddH_2O, 13.7 ul glacial HOAc. Stored at 4° C.

TE: 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8. Stored at room temp.

Methods

Each 40 ml bacterial culture was harvested in a 30 ml Corex tube by centrifugation for 10 minutes at 5000 rpm in a RC2-B automatic, refrigerated Sorval centrifuge. The pellet was

resuspended in 1 ml of GTE and transferred into a 15 ml Corex tube. After 5 minutes at room temperature, 100 ul of lysozyme was added and the mixture was incubated for 15 minutes on ice. Next 2 ml of 0.2 N NaOH/1% SDS was added. (This solution must be freshly prepared just prior to use.) The sample was then vortexed and placed on ice for 5 minutes. After centrifugation at 10,000 rpm for 5 minutes at 4° C , the supernatant was transferred to a new 15 ml corex tube and 9 ml of 95% ethanol (-20° C) was added. The supernatant and ethanol were mixed together thoroughly and placed on ice for 5 minutes. The solution was recentrifuged and the pellet was resuspened in 250 ul of TE buffer. It was then transferred to a new 1.5 ml Eppendorf tube and 125 ul of 7.5 M NH_4OAc was added. After placing it on ice for 10 minutes, it was centrifuged in a Brinkman Eppendorf centrifuge for 5 minutes at 4° C .

The supernatant was transferred to a new tube and 1ml of 95% ethanol (-20° C) was added. The contents of the tube were mixed well and placed on ice for 5 minutes. The pellet was recovered by centrifugation in 5 minutes at 4° C . Following this the pellet was dissolved in 50 ul of TE buffer and reprecipitated by adding 25 ul of 7.5 M NH_4OAc and 150 ml of 95% ethanol (-20° C). The sample was vortexed and placed on ice for 5 minutes. The pellet was recovered by centrifugation in 5 minutes. at 4° C . The pellet was resuspended in 50-100 ul of TE buffer to a final DNA concentration of 1 ug/ul.

At this point the DNA may be RNAsed by adding 2ul [10 mg/ml] boiled RNAse for 30 minutes at room temperature, phenol-chloroform extracted, reprecipitated, and resuspended in TE.

A.4 Genomic DNA Extraction

A.4.1 Genomic DNA Extraction from Multiple Flies

Genomic DNA from multiple flies for characterization of isolines was extracted by Lifton protocol, as modified by Loretta Arcangeli for Dr. Haymer's lab.

Materials

Lifton Grind buffer: 0.2M sucrose, 50 mM EDTA, 100 mM Tris pH9.0, 0.5% SDS. Add dH_2O to 100 mls. Filter sterilize.

Fiberfil strainer: A cotton ball-sized piece of sterile polyester fiberfill placed into a lOcc syringe

Proteinase K: 20 mg/ml. Stored at -20 C.

8M KAc:

RNAse A:

CHIASM:

3M NaAc pH 6.0:

95% ETOH:

TE:

Methods

Live flies were etherized just prior to grinding. Approximately 0.3 g of flies were homogenized in 5 mls of Lifton Grind buffer on ice. The homogenate was strained through a fiberfill strainer into a 15ml polypropylene tube on ice. 50 ul of 20 mg/ml Proteinase K was added and the tube was incubated at 65°C for one hour. 750ul of 8M KAc was added and the tube placed on ice at for 1 hour. At this point the tube was centrifuged for 15 minutes at 10,000 rpm 4°C in a RC2-B automatic, refrigerated Sorval centrifuge. The supernatant was transferred to another tube and ethanol precipitated with 2 volumes of room temperature 95% EtOH. The pellet was resuspended in 500ml TE and transfered to an Eppendorf tube. 2.5ul of 10 mg/ml RNAse A was added and the sample was incubated at room temperature for 30 minutes. The sample was extracted with an equal volume of buffered phenol and extracted with an equal volume of CHISAM. The aqueous layer was transferred to another tube and precipitated by adding 0.1 volume 3M NaAc pH 6.0 and 2.5 volumes of ice cold 95% EtOH. The pellet was resuspended in 0.2 to 0.5 ml TE.

A.4.2 Genomic DNA Extraction from Individual Flies

Genomic DNA from individual flies was extracted via PureGene Animal Tissue (Gentra Systems, Inc.) protocol modified for single flies.

Materials

PureGene Animal Tissue Kit

Grinding solution: 3 ul of 20 mg/ml Proteinase K to 600 ul of Cell Lysis Buffer

Diluted RNAse A: 6 ul RNase + 114 ul Cell Lysis Buffer

Methods

A single fly was added to 30 ul of grinding solution in a 1.5ml Eppendorf tube on ice. The fly was homogenized with a Kontes grinder and incubated at 50° C for 3 hours or up to overnight. 3 ul of diluted RNAse A was added and incubate was continued at 37° C for 15-60 minutes. The tube was placed on ice and 10 ul of Protein Precipitation Solution was added. The tube was centrifuged at 14K for 3-4 minutes. the supernatant was removed to a new 1.5ml Eppendorf tube containing 30 ul of 100% isopropanol and centrifuge at 14K for 1 minute. The pellet was rinsed with 30 ul of 70% EtOH and air dried for at least 15 minutes. The pellet was resuspended in 10 ul of DNA Hydration Solution.

A.5 Procedure for Squash Blots

Squash blots followed the protocol of Martin (1990).

Materials

Denaturing solution: 1.5 M NaCl; 0.5 M NaOH

Neutralizing solution: 1 M Tris, pH 8.0; 1.5 M NaCl

Prewash: 5xSSC; 0.5% SDS; 1mM EDTA

20X SSC: 175.3 g NaCl, 88.2 g Na citrate, add 800 ml ddH20, pH 7.0, and fill to 1 L. Autoclaved. (3 M NaCl, 0.3 M Na citrate)

Prehybridization solution in formamide: 12.5 ml 20X SSC, 2.5 g Blocking reagent, 0.5 ml 10% sarkosyl, 50 ul 10% SDS, 9.45 ml ddH20, 25 ml 100% formamide. Stored at -20 C in 50 ml aliquots.

Hybridization solution: Dig labelled probe in prehybridization solution.

Methods

Euthanized flies were placed on a MagnaGraph nylon filter, sandwiched between plastic sheets, and squashed using a metal spatula. The squash blot was air dried for 2 minutes. The filter was placed in denaturing solution with gentle shaking, for 5 minutes. The loose fly parts were removed with forceps while the filter was in the denaturing solution. The filter was then placed in neutralizing solution and gently shaken for 5 minutes, 2 times. The filter was rinsed briefly in 2x SSC, air dried for 30 minutes, and Vacuum bake at 80 degrees for 1 hour. The filter was then washed in 2x SSC for 5 minutes and incubated at 50° C in Prewash for 15 minutes. The remaining fly tissue was gently scraped off using a kimwipe soaked in prewash. The blot was placed in prehybridize solution overnight at 37° C and hybridized overnight at 42° C. The blot was visualized using LumiPhos Kit from Boehringer-Mannheim Corp.

A.6. Digoxigenin Labeled Probes

Plasmids PGN156 and H_0 were used to make randomly labeled digoxigenin-dUTP probes. The inserts were excised from their respective vectors by an EcoRI digest which released inserts of 7K and 2.5K, respectively. The inserts were isolated on a 1% agarose gel, gene cleaned using the Geneclean II Kit (Bio 101, Inc.), and dig-labelled according to the Genius protocol. The following protocol was obtained from the DNA labeling Kit (Genius) with slight modifications.

Materials

Provided by Genius Kit:

Unlabeled control-DNA: [200 ug/ml] pBR328 20 ul digested with BamHI. Stored at 4° C.

Hexanucleotide mixture: 80 ul of 10X concentrated hexanucleotide reaction mixture.

dNTP Labeling Mixture: 80 ul of 10X concentrated dNTP labeling mixture containing 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, 0.35 mM DIG-dUTP. pH 6.5. Stored at -20° C.

Klenow enzyme: (NEB) 2 units/ul. Stored at -20° C.

Methods

1 ug of insert from either PGN156 or H_0, purified by Geneclean (See Geneclean II Kit for details), was suspended into 15 ul of sterile distilled water. The DNA was denatured for 10 minutes at 95° C and chilled quickly on ice.

The following were added to the denatured DNA on ice: 2 ul of hexonucleotide mixture (vial 5), 2 ul dNTP mixture (vial 6), and 1 ul Klenow enzyme (NEB).

The control-DNA (vial 2) was labeled in the same way. The reaction was incubated at 37° C overnight and then purified using the Geneclean kit.

A.7 Southern Blot

A.7.1 Denaturing, Neutralizing, and Blotting the Gel

Materials

10X TBE: 108 g Tris base, 55 g boric acid, 40 ml 0.5M EDTA (pH8), brought up to 1 L with ddH_2O. Stored at room temp.

Dig-labelled Lambda Hind III: [0.3 ug/ul] Produces 8 fragments with base pairs as follows: 23130, 9416, 6557, 4361, 2322, 2027, 564, and 125. Stored at 4° C.

Denaturing solution: 1.5 M NaCl, 0.5 M NaOH (87.6 g/L NaCl, 20g/L NaOH)

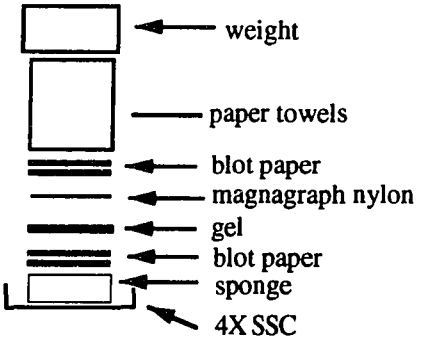
Neutralizing solution: 1 M Tris pH 8.0 (Tris base 5.3 g/L, Tris acid 8.88 g/L), 1.5 M NaCl (87.6 g/L)

Methods

Single fly extracted genomic DNA digested with the appropiate enzyme were electrophoresised in a 0.7% agarose gel in 1X TBE with ethidium bromide against Dig-labelled Lambda Hind III overnight at

40V. In the morning the gel was denatured for 1 hour in denaturing solution at room temperature. Next the gel was neutralized for 1 to 1.5 hour in neutralizing solution at room temperature.

The Southern blot was set up as follows:



The MagnaGraph was cut to fit the size of the gel. The gel should be bordered with parafilm so that none of the blot paper below the gel touches any of the blot paper above the gel. The blot sat overnight or all day.

A.7.2 Pre-Hybridization and Hybridization of the Membrane

Materials

20X SSC: 175.3 g NaCl, 88.2 g Na citrate, add 800 ml ddH20, pH 7.0, and fill to 1 L. Autoclaved. (3 M NaCl, 0.3 M Na citrate)

Prehybridization solution in formamide: 125 ml 20X SSC, 25g Blocking reagent, 5ml 10% Sarkosyl, 0.5ml 10% SDS, 94.5ml water, 250ml100% Formamide per 0.5L. 50ml aliquots were stored at -20 C.

of room temperature 95% EtOH. The pellet was resuspended in 500ml TE and transfered to an Eppendorf tube. 2.5ul of 10 mg/ml RNAse A was added and the sample was incubated at room temperature for 30 minutes. The sample was extracted with an equal volume of buffered phenol and extracted with an equal volume of CHISAM. The aqueous layer was transferred to another tube and precipitated by adding 0.1 volume 3M NaAc pH 6.0 and 2.5 volumes of ice cold 95% EtOH. The pellet was resuspended in 0.2 to 0.5 ml TE.

A.4.2 Genomic DNA Extraction from individual flies

Genomic DNA from individual flies was extracted via PureGene Animal Tissue (Gentra Systems, Inc.) protocol modified for single flies.

Materials

PureGene Animal Tissue Kit

Grinding solution: 3 ul of 20 mg/ml Proteinase K to 600 ul of Cell Lysis Buffer

Diluted RNAse A: 6 ul RNase + 114 ul Cell Lysis Buffer

Methods

A single fly was added to 30 ul of grinding solution in a 1.5ml Eppendorf tube on ice. The fly was homogenized with a Kontes grinder and incubated at 50° C for 3 hours or up to overnight. 3 ul of diluted RNAse A was added and incubate was continued at 37° C for 15-60 minutes. The tube was placed on ice and 10 ul of Protein Precipitation Solution was added. The tube was centrifuged at 14K for 3-4 minutes. the supernatant was removed to a new 1.5ml

Methods

Before starting the post hybridization washes, 50mls of Buffer B (per 15cm x 20cm blot) were made fresh and 20 mls put aside for the antibody conjugate reaction.

The hybridization solution was poured off and saved in a sterile 50 ml tube, and stored it at -20 C for up to 3 months.

Washed the filter 2 times at room temperature in 2X SSC; 0.1% SDS for 5 minutes. Washed the filter 1 time at 68° C in 0.1X SSC; 0.1% SDS for 15 minutes. Washed the filter 1 time at 68° C in 0.5X SSC; 0.1% SDS for 15 minutes.

After the post-hybridization washes, the membrane was equilibrated in Buffer A for 1 minute and blocked in 30 mls of Buffer B for 3 hours with gentle shaking.

A.7.4 Lumiphos Detection Procedure

Materials:

Anti-Digognigenin alkaline phosphatase conjugate: vial 8 of the Genius Nonradioactive DNA Labeling and Detection Kit

Lumi-phos

X-ray film

Methods

Near the end of the blocking step, the anti-digoxigenin alkaline phosphatase conjugate was diluted to 1:5000 in Buffer B for a working concentration of 150 mU/ml. The membrane was removed

from Buffer B and transferred to the antibody conjugate solution. It was then incubated for 30 minutes at room temperature with gentle shaking .

The membrane was then washed twice for 15 minutes each in Buffer A. The washed membrane was equilibrated in Buffer C for 2 minutes to prepare it for visualization.

Prior to visualiztion Lumi-Phos was prewarmed to room temperature (2mls per 15cm x 20 cm membrane).

The membrane was then carefully transferred to a new hybridization bag from Buffer C and the Lumi-Phos solution was added. The membrane was incubated in the dark for 1 minute at room temperature. Excess liquid was removed and the hybridization bag sealed.

The membrane was incubated at 37° C for 15-30 minutes in the dark before being exposed to X-ray film *(e.g.,* Kodak XAR film in a darkroom. The initial recommended exposure time is 15 minutes. After the 30 minute/+37° C incubation, the light emission remains constant for 24 hours.

Sd genotype and Responder genotype data were collected from these blots.

References

Andersson, S., and A. Lambertsson. 1993. Evolution of the dec-1 Eggshell locus in *Drosophila*. II. Intraspecific DNA sequence analysis reveals length mutations in a repetitive region in *D. melanogaster*. *Journal of Molecular Evolution* 36(6): 536-544.

Anxolagehere, Dominique, Danielle Nouaud, Georger Periquet and Paul Tchen. 1985, P-element distribution in Eurasian populations of *Drosophila melanogaster*: a genetic and molecular analysis. *Proceedings of the National Acadamy of Science, USA* 82: 5418-5422.

Ashburner, Michael. 1989. *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Press.

Aslanidis, C., G. Jansen, C. Amemiya, G. Shytler, M. Mahadevan, C. Tsifidia, C. Chen, J. Alleman, N. G. M. Wromskamp, M. Voojis, J. Buxton, K. Johnson, H. J. M. Smeets, G. G. Lennon, A. V. Carrano, R. G. Korneluk, B. Wieringa, and P. J. de Jong. 1992. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 355: 548-551.

Bengtsson B. O. and M. K. Uyenoyama. 1990. Evolution of the segregation ratio: modification of gene conversion and meiotic drive. *Theoretical Population Biology* 38(2):192-218.

Berg, D. E. and M. M. Howe. 1989. *Mobile DNA*. Am. Soc. Microbial., Washington, D. C.

Beyer, William H. (Ed).1968. *Handbook of Tables for Probability and Statistics*. CRC Press. Cleveland Ohio.

Biancalana, V., F. Serville, J. Pommier, J. Julien, A. Hanauer and J. L. Manndel. 1992. Moderate instability of the trinucleotide repeat in spino-bulbar atrophy. *Human Molecular Genetics* 1: 255-258.

Brittnacher, J. G. and B. Ganetsky. 1983. On the components of Segregation Distortion in *Drosophila melanogaster*. II. Deletion mapping and dosage analysis of the *Sd* locus. *Genetics* 103:659-673.

Buxton, J., P. Shelbourne, J. Davies, C. Jones, T. Van Tongeren, C. Aslanidis, P. de Jong, G. Jansen, M. Anvret, B. Riley, R. Williamson and K. Johnson. 1992. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 335: 547-548.

Cabot, Eric L., Parul Doshi, Ma-Lien Wu and Chung-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, Brian, and Daniel Hartl. 1978 Population Dynamics of the Segregation Distorter polymorphism of *Drosophila melanogaster*. Genetics 89: 171-192.

Charlesworth, B., C. H.Langley and W. Stephan. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112: 947-962.

Costa, R., A. A. Peixoto, J. R. Thackeray, R. Dalgleish and C. P. Kyriacou. 1991. Length polymorphism in the threonine-glycine-encoding repeat region of the period gene in *Drosophila*. *Journal of Molecular Evolution* 32: 238-246.

Crow, James F. 1979. Genes that violate Mendel's Rules. *Scientific American* 240: 134-146.

Crow, James F. 1988. The Ultraselfish Gene. *Genetics* 118: 389-391.

Crow, James F. 1991. Why is Mendelian segregation so exact? *BioEssays* 13: 1-8.

Crow, James F., C. Thomas and L. Sandler. 1962. Evidence that the segregation-distortion phonomenon in *Drosophila* involves chromosome breakage. *Proceedings of the National Acadamy of Science, USA* 48: 1307-1314.

Current Protocols in Molecular Biology. 1987. ed. Frederick M. Ausubel. John Wiley and Sons. New York.

Deville, P., P. Slagboom, C. J. Cornelisse and P. L. Pearson. 1986. Permanent attachment of replication origins to the nuclear matrix in BHK-cell. *Nucleic Acids Research* 14: 3241-3249.

Dominguez, A., E. Santiago, J. Albornoz and A. Gutierrez, A. 1993. The Segregation Distorter (SD) complex and the accumulation of deleterious genes in laboratory strains of *Drosophila melanogaster*. *Theoretical and Applied Genetics* 87: 479-486.

Dover, Gabriel A., 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.

Doshi, Parul, Sunjay Kaushal, Cheeptip Benyajati and Chung-I Wu. 1991. Molecular analysis of the Responder satellite DNA in *Drosophila melanogaster*: DNA bending, nucleosome structure and Rsp-binding proteins. Molecular Biology and Evolution. 8(5): 721-741.

Dover, Gabriel A., 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.

Dover, Gabriel. 1995. Slippery DNA runs on and on and on... *Nature Genetics* 10: 254-256.

Durfy, S. J., and H. F. Willard. 1989. Patterns of intra- and interarray sequence variation in alpha satellite form the human X chromosome: evidence for short range homogenization of tandemly repeated DNA sequences. *Genomics* 5: 810-821.

Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice*. Sage Press. Newbury Park, CA.

Engels, William R. 1989. P-elements in *Drosophila melanogaster*. pp. 437-484. In: *Mobile DNA*. ed. Berg, D. E. and M. M. Howe. American Society for Microbiology, Washington, D. C.

Engels, William R., Dena M. Johnson-Schlitz, William B. Eggleston and John Sved. 1990. High frequency P element loss in *Drosophila* is homolog dependent. *Cell* 62: 515-525.

Eshel, Ilan. 1984. Are intra gametic conflicts common in nature? are they important to evolution? *Journal of Theoretical Biology* 198:77-80.

Eshel, Ilan. 1985. Evolutionary genetic stability of Mendelian segregation and the role of free recombination in the chromosomal system. *American Naturalist* 125(3): 412-420.

Eshel, Ilan. and M. W. Feldman. 1982. On the evolutionary genetic stability of the sex ratio. *Theoretical Population Biology*. 21: 430-439.

Feldman, Marcus W., and Sarah P. Otto. 1991. A comparative approach to the population genetic theory of segregation distortion. *The American Naturalist* 137(3): 443-456.

Felsenstein, J. 1974. The evolutionary advantage of recombination. . *Genetics* 78: 737-756.

Fu, Y. H., D. P. A. Kuhl, A. Pizzuti, M. Pieretti, J. S. Sutcliffe, S. Richards, A. J. M. H. Verkerk, J. J. A. Holden, R. G. J. Fenwick, S. T. Warren, B. A. Oostra, D. L. Nelson and C. T. Caskey. 1991. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67: 1047-1058.

Ganetzky, B. 1977. On the components of segregation distortion in *Drosophila melanogaster. Genetics* 86: 321-355.

Gatti, M., G. Santini, S. Pimpinelli and G. Olivieri. 1979. Lack of spontaneous sister chromatid exchanges in somatic cells of *Drosophila melanogaster. Genetics* 91: 255-274.

Gennarelli, M., B. Dallapiccola, M. Baiget, L. Martorell and G. Novelli. 1994. Meiotic drive at the myotonic dystrophy locus. *J. Med. Genet.* 31: 980.

Gloor, G. B., N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston and W. R. Engels, 1991. Targeted gene replacement in *Drosophila* via P element-induced gap repair. *Science* 253:1110-1117.

Gray, I. C. and A. J. Jefferys. 1991. Evolutionary transience of hypervariable minisatellites in man and the primates. *Proceedings of the Royal Society of London Series B* 243:241-253.

Groden, J., and J. German. 1992. Bloom's syndrome XVIII. Hypermutability at a tandem-repeat locus. *Human Genetics* 90: 360-367.

Haig, D. and A. Grafen. 1991. Genetic scrambling as a defense against meiotic drive. *Journal of Theoretical Biology* 153 (4): 531.

Hancock, John M. and Gabriel A. Dover, 1988. Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/29S r RNAs. *Mol. Biol. Evol.* 5(4): 377-391.

Harding, R. M. 1992. The evolution of tandemly repetitive DNA: recombination rules. *Genetics* 132(3): 847.

Harley, H., S. Rundle, S. Crows, W. Reardon, A. Buckler, P. Harper, D. Houseman and D. Shaw. 1992. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 355: 545-546.

Hartl, Daniel. 1970. Meiotic drive in natural populations of *Drosophila melanogaster*. IX. Suppressors of Segregation Distorter in wild populations. *Canadian Journal of Genetics and Cytology*. 12: 594-600.

Hartl, Daniel. 1972. Population dynamics of sperm and pollen killers. *Theoretical and Applied Genetics*. 42:81-88.

Hartl, Daniel. 1973. Complementation analysis of male fertility among the Segregation Distorter chromosomes of *Drosophila melanogaster*. *Genetics* 73: 613-629.

Hartl, Daniel. 1977. Mechanisms of a case of genetic coadaptation in *Drosophila melanogaster*. *Proceedings of the National Acadamy of Science, USA* . 74: 324-328.

Hartl, D. and N. Hartung. 1975. High frequency of one element of segregation distorter in natural populations of *Drosophila melanogaster*. *Evolution* 29: 512-518.

Hartl, Daniel and Yuichiro Hiraizumi. 1976. Segregation Distortion. In: Ashburner, M. and E. Novitski (Eds.) 1976. *The Genetics and Biology of Drosophila* 1b: 615-666.

Hickey, Donald A., Ada Loverre and George Carmody. 1986. Is the segregation distortion phenomenon in *Drosophila* due to recurrent active genetic transposition? *Genetics* 114: 665-668.

Hiraizumi, Y. 1961. Lethality and low viability induced by the segregation distorter locus (symbol SD) in *Drosophila melanogaster. Annual Report of the National Instutite of Genetics of Japan.* 12: 1-2.

Hiraizumi, Y. and James F. Crow. 1960. Heterozygous effects on viability, fertility, rate of development, and longevity of *Drosophila* chromosomes that are lethal when homozygous. *Genetics* 45: 1071-1084.

Hiraizumi, Y. and Anita M. Thomas. 1984. Suppressor systems of Segregation Distorter (SD) chromosomes in natural populations of *Drosophila melanogaster. Genetics* 106 (2): 279-292.

Hiraizumi, Y., Sandler, L., and James F. Crow. 1960. Meiotic drive in natural populations of *Drosophila melanogaster.* III. Populational implications of the Segregation Distorter locus. *Evolution* 14: 433-444.

Hurst, Laurence D. and Andrew Pomiankowski. 1991. Causes of sex ratio bias may account for unisex sterility in hybrids: A new explanation of Haldane's rule and related phenomena. *Genetics* 129:841-858.

Ikeuchi, T., S. Igarashi, Y. Takiyama, O. Onodera, M. Oyake, H. Takano, R. Koide, H. Tanaka and S. Tsuji. 1995. Non-Mendelian transmission in Dentatorubral-Pallidoluysian Atrophy (DRPLA) and Machado-Joseph Disease (MJD): mutant allele is preferentially transmitted in male meiosis. *Am. J. Hum. Genet.* (In Press).

Jefferys, A. J., V. Wilson and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* 314: 67-73.

Jefferys, A. J., N. J. Royle, V. Wilson and Z. Wong. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332: 278-281.

Jefferys, A. J., A. MacLeod, K. Tamaki, D. L. Neh and D. G. Monckton. 1991. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 254: 204-209.

Jefferys, A. J., Z. Wong, V. Wilson, I Patel, R. Neumann, N. Royle and J. A. L. Armour. 1989. Applications of multilocus and single-locus minisatellite DNA probes in forensic medicine. In: Ballantyne, Jack, George Sensabaugh and Jan Witkowski. *DNA Technology and Forensic Science. Banbury Report #32.* Cold Spring Harbor Laboratory Press.

Jones, R. S., and S. S. Potter. 1985. Characterization of cloned human alphoid satellite with an unusual monomeric construction: evidence for enrichment in HeLa small polydispersed circular DNA. *Nucleic Acids Research* 13: 1027-1042.

Kataoka, Y. 1967. A genetic system modifying segregation distortion in a natural population of *Drosophila melangaster* in Japan. *Japanese Journal of Genetics.* 42: 327-337.

Kauffman, S. A. 1993. *The Origins of Order - Self-Organization and Selection in Evolution.* New York: Oxford University Press.

Klein J., P. Sipos and F. Figueroa. 1984. Polymorphism of *t*-complex genes in European wild mice. *Genetic Research (Cambridge)* 44:39-46.

Klitz, William, Wing Kai Lo, Meinhard Neugebauer, Max P. Baur, Ekkehard D. Albert and Glenys Thomson. 1987. A comprehensive search for segregation distortion in HLA. Human Immunology. 18: 163-180.

Kruger, J., and F. Vogel. 1975. Population genetics of unequal crossing over. *Journal of Molecular Evolution* 4: 210-247.

Laspada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding and H. Fishbeck. 1991. Androgen receptor gene mutations in Z-linked spinal and bulbar atrophy. *Nature* 352: 77-79.

Lemeunier, Francoise and Sylvie Aulard. 1992 Inversion Polymorphism in *Drosophila melanogaster*. In: Krimbas, Costas B. and Jeffrey R. Powell (Eds.) *Drosophila* Inversion Polymorphisms. CRC Press. Boca Raton, FL.

Lennington, Sarah and Kathleen Egid. 1989. Environmental influences on the preferences of wild female house mice for males of differing *t*-complex genotypes. *Behavior Genetics* 19 (2): 257-263.

Lennington, S., Franks, P., and J. William. 1988. Distribution of *t*-haplotypes in natural populations of wild house mice. *Journal of Mammology* 69:489-499.

Lessard, Sabin. 1985. The role of recombination in the modifier theory of autosomal segregation distortion. *Theoretical Population Biology* 28: 133-149.

Levinson, G. and G. A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

Liberman, Uri. 1976. Modifier theory of meiotic drive: Is Mendelian segregation stable?. *Theoretical Population Biology* 10: 127-132.

Liberman, Uri. 1990. Mendelian segregation: the choice between "order" and "chaos". *Journal of Mathematical Biology* 28: 435-449.

Liberman, Uri, and Marcus W. Feldman. 1982. On the evolution of fluctuating segregation distortion. *Theoretical Population Biology* 21: 301-317.

Lyttle, Terrence W. 1979. Experimental population genetics of meiotic drive systems. II. Accumulation of genetic modifiers of segregation distorter (SD) in laboratory populations. *Genetics* 91: 339-357.

Lyttle, Terrence W. 1986. Additive effects of multiple Segregation Distorter (SD) chromosomes on sperm dysfunction in *Drosophila melanogaster. Genetics* 114: 203-216.

Lyttle, Terrence W. 1991. Segregation distorters. Annual Review of *Genetics* 25: 511-557.

Lyttle, Terrence W., 1993. Cheaters sometimes proper: distortion of Mendelian segregation by meiotic drive. *Trends in Genetics* 9(6):205-210.

MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexier, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., Kirby, S., North, M., Youngman, S., Mott, R., Zehetner, G., Sedlacek, Z., Poustka, A., Frischauf, A.-M., Lahrach, H., Buckler, A. J., Church, D., Doucette-Stamm, L., O'Donovan, M. C., Riba-Ramirez, L., Shah, M., Stanton, V. P., Strobel, S. A., Draths, K. M., Wales, J. L., Dervan, P., Housman, D. E., Alther, M., Shiang, R., Thompson, L., Fielder, T., Wasmuth, J. J., Tagle, D., Valdes, J., Elmer, L., Allard, M., Castilla, L., Swaroop, M., Blanchard, K., Collins, F. S., Snell, R., Holloway, T., Gillespie, K., Datson, N., Shaw, D., and Harper, P. S. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes. *Cell* 72: 971-983.

Martin, Diana W. and Yuichiro Hiraizumi. 1979. On the models of segregation distortion in *Drosophila melanogaster. Genetics* 93: 423-435.

Martin, Jayma L. 1990. Untitled Masters project. Department of Genetics, University of Hawaii.

McClean, Janna R., Cynthia J. Merrill, Patricia A. Powers and Barry Ganetzky. 1994 Functional identification of the Segregation distorter locus of *Drosophila melanogaster* by germline transformation. *Genetics* 137: 210-209.

McMurray, Cynthia T. 1995. Mechanisms of DNA expansion. *Chromosoma* 104:2-13.

Miyazaki, W. Y. and T. L. Orr-Weaver. 1992. Sister-Chromatid Misbehavior in *Drosophila ord* Mutants. *Genetics* 342: 1047-1061.

Mukai, T. 1964. The genetic structure of natural populations of *Drosophila melanogaster.* I Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50: 1-19.

Muller, H. J. 1964. The relation of recombination to mutational advance. *Mutational Research.* 1: 2-9.

Newfeld, S. J., A. T. Schmid and B. Yedvobnick. 1993. Homopolymer length variation in the *Drosophila* gene mastermind. *Journal of Molecular Evolution* 37: 483-495.

Newfeld, S. J., H. Tachida and B. Yedvobnick. 1994. Drive-selection equilibrium: homopolymer evolution in the *Drosophila* gene mastermind. Journal of Molecular *Evolution* 38: 637-641.

Oakeshott, J.G, J. B. Gibson, P. R. Anderson, W. R. Knibb, D. G. Anderson and G. K. Chambers. 1981. Alcohol dehydrogenase and glycerol-3 phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution* 36: 86-96.

Oakeshott, J. G., B. K. Chambers, J. B Gibson, W. F. Eanes and D. A. Willcocks. 1983. Geographic variation in *G6pd* and *Pgd* allele frequencies in *Drosophila melanogaster.* Heredity 50: 67-72.

Ohta, T. 1980. *Evolution and Variation of Multigene Families. (Vol. 37).* Berlin: Springer-Verlag.

Ohta, T. 1981. Genetic variation in small multigene families. *Genetical Research* 37: 133-149.

Ohta, T. 1983. Theoretical study on the accumulation of selfish DNA. *Genetical Research* 41: 1-15.

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(1): 209.

Perelson, A. S., and G. I. Bell. 1977. Mathematical models for the evolution of multigene families by unequal crossing over. *Nature* 265: 304-310.

Periquet, G. 1979. Maintenance of the 'atrophie gonadique' character in *Drosophila melanogaster* : theoretical models for natural and experimental evolution. Genetica. 51: 33-43.

Peterson, H. M. and J. R. Laughnan, 1963. Intrachromosomal exchanges at the Bar locus in *Drosophila. Proceedings of the National Acadamy of Science, USA* 50:126-133.

Powers, Patricia and Barry Ganetsky. 1991. On the components of Segregation Distortion in *Drosophila melanogaster*. V. Molecular analysis of the *Sd* locus. *Genetics* 129:133-144.

Prout, T., J. Bundgaard and S. Bryant. 1973. Population genetics of modifiers of meiotic drive. I: the solution of a special case and some general implications. *Theoretical Population Biology* 4: 446-465.

Ray, J. H. and J. German. 1983. The cytogenetics of the "chromosome-breakage syndromes." In: German, J (ed.) *Chromosome mutation and neplasia*. (pp. 135-167). A. R. Liss, New York.

Ritossa, F. 1976. The Bobbed Locus. In: Ashburner, M. and E. Novitski (Eds.) 1976. *The Genetics and Biology of Drosophila* 1b: 801-846.

Roberts, Paul A. 1976. The Genetics of Chromosome Aberration. In: Ashburner, M. and E. Novitski (Eds.) 1976. *The Genetics and Biology of Drosophila* 1a: 67-184.

Rubinsztein, D. C., W. Amos, J. Leggo, S. Goodburn, R. S. Ramesar, J. Old, R. Bontrop, R. McMahon, D. E. Barton and M. A. Ferguson-Smith. 1994. Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat. Genet.* 7(4): 525-530.

Rubinsztein, David C., William Amos, Jayne Leggo, Sandy Goodburn, Sanjeev Jain, Shi-Hua Li, Russell, L. Margolis, Christopher A. Ross and Malcolm A. Ferguson-Smith. 1995. Microsatellite evolution - evidence for directionality and variation in rate between species. *Nat. Genet.* 10: 337-343.

Sandler, L. and Kent Golic. 1985. Segregation Distortion in *Drosophila. Trends in Genetics.* 1: 181-185.

Schlotterer, C., and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20: 211-215.

Sharp, Cecil B., Arthur J. Hilliker and David G. Holm. 1985. Further characterization of genetic elements associated with the Segregation Distorter phenomenon in *Drosophila melanogaster. Genetics* 110: 671-688.

Shimakawa, E. S. 1987. A molecular and population genetic study of Segregation distorter in *Drosophila melanogaster.* Ph.D. dissertation, University of Hawaii, Manoa

Silver, Lee M. 1985. Mouse *t* haplotypes. *Annual Review of Genetics* 19: 179-208.

Smith, G. P. 1974. Unequal crossover and the evolution of multigene families. *Proceedings of the Cold Spring Harbor Symposium on Quantitative Biology* 38: 507-513.

Smith, George P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-535.

Smoller, D., C. Friedel, A. Schmid, D. Bettler, L. Lam and B. Yedvobnick. 1990. The *Drosophila* neurogenic locus mastermind encodes a nuclear protein unusually rich in amino acid homopolymers. *Genes and Development* 4: 1688-1700.

Stalker, H. D. 1961. The genetic systems modifying meiotic drive in *Drosophila* papamelanica. *Genetics* 46: 177-202.

Stephan, W. 1986. Recombination and the evolution of satellite DNA. *Genetical Research* 47: 167-174.

Stephan, W. 1987. Quantitative variation and chromosomal location of satellite DNA's. *Genetical Research* 50: 41-52.

Stephan W. 1989. Tandem-repetitive noncoding DNA: forms and forces. *Molecular Biol. Evol.* 6: 198-212.

Stephan, W. and S. Cho. 1994. Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* 136(1): 333-341.

Strand, M., T. A. Prolla and R. M. Liskay. 1993. Destablization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365(6443): 274.

Sved, J. A. and C. D. Laird. 1990. Population genetic consequences of the fragile-X syndrome, based on the X-inactivation imprinting model. American journal of Human Genetics. 46(3): 443-451.

Tachida, H. and M. Itzuka. 1992. Persistence of repeated sequences that evolve by replication slippage. *Genetics* 131: 479-491.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17(16): 6463-6471.

Temin, Rayla Greenberg. 1991. The independent distorting ability of the *Enhancer of Segregation Distortion, E(SD),* in *Drosophila melanogaster. Genetics* 128: 339-356.

Temin, Rayla Greenberg and Marta Marthas. 1984. Factors influencing the effect of segregation distortion in natural populations of *Drosophila melanogaster. Genetics* 107:375-393.

Temin, R., B. Ganetzky, P. A. Powers, T. W. Lyttle, S. Pimpinelli, P. Dimitri, C.-I. Wu and Y. Hiraizumi. 1991. Segregation distorter (SD) in *Drosophila melanogaster*: Genetic and molecular analyses. *The American Naturalist* 137: 287.

Thomson, Glenys J. and Marcus W. Feldman 1974. Population genetics of modifiers of meiotic drive. II. Linkage modification in the segregation distorter system. *Theoretical Population Biology* 5: 155-162.

Thomson, Glenys J. and Marcus W. Feldman. 1976. Population genetics of modifiers of meiotic drive. III. Equilibrium analysis of a general model for the genetic control of segregation distortion. *Theoretical Population Biology* 10: 10-25.

Thompson-Stewart, Dianne, Gary H. Karpen, and Allan C. Spradling. 1994. A transposable element can drive the concerted evolution of tandemly repetitious DNA. *Proceedings of the National Academy of Science, USA*. 91: 9042-9046.

Trippa, G., A. Loverre and R. Cicchetti. 1980. Cytological analysis of an SD chromosome from a natural population of *Drosophila melanogaster. Genetics* 95: 399-412.

Trippa, G. and A. Loverre. 1975. A factor on a wild third chromosome (IIIRa) that modifies the Segregation distortion phenomenon in *Drosophila melanogaster. Genetic Research (Cambridge)* 26:113-125

Trippa, G., R. Cicchetti, A. Loverre and A. Micheli. 1981. Biological parameters and the segregation distortion (SD) phenomenon in *Drosophila melanogaster. Genetics* 97: 563-580.

Valdes, A. M., M. Slatkin and N. B. Freimer. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, 133, 737-749.

Vollmer, S. J. and Charles Yanofsky. 1986. Efficient cloning of genes of *Neurospora crassa. Proceedings of the National Acadamy of Science, USA* . 83, 4869-4873.

Walsh, J. B. 1987. Persistence of tandem arrays: implications for satellite and simple sequence DNAs. *Genetics* 115: 553-567.

Wu, Chung-I and M. F. Hammer. 1990. Molecular evolution of ultraselfish genes of meiotic drive systems. In: *Evolution at the Molecular Level*, ed. R. K. Selander, T. Whittam and A. Clark. Sinauer Press.

Wu, Chung-I, John R. True and Norman Johnson. 1989. Fitness reduction associated with the deletion of a satellite DNA array. *Nature* 341: 248-251.

Wu, Chung-I, Terrence W. Lyttle, 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.

Yasuda, Norikazu and Motoo Kimura. 1968. A gene-counting method of maximum likelihood for estimation gene frequencies in ABO and ABO-like systems. *Annals of Human Genetics (London)* 31:409-420.