DIMACS Technical Report 98-51
January 1999

Genetic Drift in Populations of Ultraselfish Gene Complexes[1]

by

D. Durand[2,3] E. Bendix,[4] K. Ardlie,[5] W. Ewens,[6] L. M. Silver[7]

[1]Submitted to Genetics

[2]Permanent Member
[3]Computational Biology Group, University of Pennsylvania, Philadelphia, PA 19104-6228 and DIMACS, Rutgers University, Piscataway, NJ 08855-1179
[4]Computational Biology Group, University of Pennsylvania, Philadelphia, PA 19104-6228
[5]Center for Genome Research, Whitehead Institute, Cambridge, MA 02139
[6]Computational Biology Group, University of Pennsylvania, Philadelphia, PA 19104-6228
[7]Department of Molecular Biology, Princeton University Princeton, NJ 08544

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, Bellcore and NEC Research Institute.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

We use stochastic population models to study the evolution of Ultraselfish Gene Complexes (USGC's). USGC's are chromosomal regions characterized by segregation distortion: a heterozygote bearing the USGC passes it to more than 50% of offspring. USGC-bearing homozygotes are sterile. USGC's promote themselves at the expense of other genes in the same genome. They have been observed in animal, plant and fungal species and may lie, undetected, in many others. While the molecular drive mechanisms differ, USGC's exhibit similar genetic features, suggesting that the fundamental evolutionary mechanisms that allow their emergence may be shared.

We present Markov models and Monte Carlo simulations of genetic drift in populations of USGC's. Genetic drift, the stochastic behavior of allele frequencies in small populations, is a fundamental force in the process of evolution, yet the role of genetic drift in the evolution of USGC's is not well understood. Our analysis shows how genetic drift causes USGC's to evolve different characteristics in species typified by small and large populations, respectively. We apply our results to two well studied USGC's: the t-haplotype in *Mus musculus* and Segregation Distorter in *Drosophila melanogaster*.

1 Introduction

Ultraselfish Gene Complexes or *USGC*s are chromosomal regions that "enhance their own transmission relative to the rest of an individual's genome" [56] and "actively interfere with the functions of other genes in the same nucleus" [57]. USGC's are characterized by segregation distortion, also referred to as *transmission ratio distortion* or *TRD*. TRD occurs when a heterozygote bearing the USGC passes it to more than 50% of offspring at the expense of the wild type chromosome. A distorting gene complex with no corresponding negative selective pressure will fix rapidly in the population and, therefore, not be observable [36, 35]. Thus we expect that as additional autosomal USGC's are discovered, many will be associated with homozygous sterility or lethality.

Ultraselfish Gene Complexes are an intriguing example of selection occurring at multiple levels. They represent both competition between genes within the same genome, and between the genome and the individual that carries them. USGC's have been observed in a number of animal, plant and fungal species and may lie, undetected, in numerous other species. Although the underlying molecular mechanisms of distorted transmission differ, these ultraselfish gene complexes exhibit similar genetic features [36]. This suggests that the population genetic forces that allow such gene complexes to evolve may be shared in common. Once an ultraselfish entity arises, what properties allow it to persist in populations?

Genetic drift, the stochastic behavior of allele frequencies in small populations, is a fundamental force in the process of evolution. The mathematics of genetic drift and the corresponding evolutionary implications have been studied in great detail since the early decades of this century (see, for example, [16]). However, there have been few studies of genetic drift in populations of ultraselfish gene complexes and the role of genetic drift in the evolution of USGC's is not well understood.

In the current paper, we study the stochastic behavior of USGC's in finite, unstructured populations as a function of population size and the degree of distortion. Although our models focus on the t-haplotype, an USGC in mice, the results are quite general and can be applied to other USGC's, notably Segregation Distorter (*SD*) in *Drosophila melanogaster*. We use our results to propose novel solutions to two long-standing problems that have stymied biologists studying the t-haplotype for over forty years: why the transmission ratio seen in wild mouse populations is closer to 0.9 than 1.0 and why recessive, lethal mutations are a recent, but widespread innovation in t-haplotype evolution? Our analysis also illustrates ways in which USGC's that evolve in species with small effective population sizes (e.g., the t-haplotype in mice) differ from USGC's in large, panmictic populations (e.g., *SD* in *Drosophila*).

In addition, our results are useful for interpreting complex modeling efforts, in which genetic drift in subpopulations has an impact on the dynamics of a larger structured population. They also suggest how measurements should be performed in the presence of stochastic variation in simulations of populations of USGC's.

In the next section, we review the biology of USGC's and present our abstract model of these systems. A Markov model of genetic drift in populations of USGC's is introduced in Section 3. In Section 4, the results of Monte Carlo simulations of the Markov model are presented. These show how the persistence of an USGC depends on distortion levels and population size. Previous modeling efforts are reviewed in Section 5 and the results of the current paper are discussed in light of this earlier work. The evolutionary implications of our work are discussed in Section 6.

2 A Model of Ultraselfish Gene Complexes

Ultraselfish gene complexes have been discovered in mice [49, 50], Neurospora [53], the tomato [44], wheat [33], Podospora [41] and many species of Drosophila [5, 21, 25]. Since the lack of a specific phenotypic manifestation of USGC's makes them difficult to detect, the fact that they have been observed in five of the twenty to thirty model organisms ever subjected to detailed genetic analysis suggests that undiscovered USGC's may exist in many other species. Although the underlying physiological mechanisms of ultraselfish behavior differ from species to species, on the genetic level several common features have been found in all systems that have been well studied: diploidy, multiple linked genes, transmission ratio distortion and homozygous male infertility [35, 36]. The most well studied USGC's are the t-haplotype in mice and *SD* in *Drosophila*. We present a model of an USGC that is based on t but can be applied to both systems.

2.1 The Biology of the t-haplotype

One of the best understood examples of USGC's is the t-haplotype in mice, a variant form of a 30 megabasepair region on chromosome 17 that is able to propagate itself through the male germ line at the expense of the wild type form of this region found on its partner chromosome [49, 50]. t-haplotypes are distinguished from their wild type counterparts by a number of characteristics. First, a series of inversions in the region results in a localized suppression of recombination and thus allows a complete t-haplotype to be transmitted as a unit from one generation to the next. Second is the preferential transmission of t-bearing chromosomes from heterozygous ($+/t$) males to 90% of their offspring, on average [3, 11]. We designate the fraction of offspring of heterozygous males that receive the t-haplotype by the variable, τ. A third property expressed by most naturally-occurring t-haplotypes is a recessive embryonic lethality. Finally, t/t males homozygous for non-lethal t-haplotypes are unconditionally sterile. The genetic data suggest that the same genes involved in TRD are also responsible for the homozygous sterility [34].

2.2 Size and Structure of Mouse Populations

There is evidence of substructuring on several levels in wild mice populations. Various studies have found evidence, both physical and genetic, that mouse populations are subdivided among different farms or villages [1, 48, 32, 10]. Evidence based on variations in gene frequency over short time scales (several generations) suggests that the effective population size of these units is on the order of 10-40 individuals. Gene frequencies over longer time scales indicate that local populations are linked by migration yielding larger effective population sizes [10]. A few studies [1, 2, 48] indicate that some of the very large populations might be further subdivided into much smaller "demes" that are true territorial/social units.

2.3 Phylogenetic Studies of the t-haplotype

Phylogenetic studies of the evolution of the t-haplotype suggests that it evolved in a step-wise manner through the sequential accumulation of four inversions [19, 38, 50]. Because each inversion is associated with loci involved in the TRD phenotype, TRD probably increased with each additional inversion over the period from three million years ago to about 100,000 years ago. Thus, we assume that TRD evolved slowly from $\tau = 0.5$ to its present level. Despite this antiquity, sequence comparisons among modern t-haplotypes show almost no nucleotide polymorphism in comparison to wild type chromosomes, suggesting that all contemporary t-haplotypes share a much more recent common ancestor that lived between 10,000 and 100,000 years ago. In stark contrast to the lack of polymorphism among independent t-haplotypes, there are a large number of independent recessive lethal mutations, which suggests both that the lethals are a recent accumulation and that they confer a selective advantage on the haplotypes that carry them. Additionally, the existence of many different lethal variants supports the theory that mouse populations are subdivided and not panmictic.

2.4 A Model of USGC's

We present a model of a population of N diploid individuals. Two alleles are represented in this population: a wild type (+) and an USGC (t). Although USGC's are known to contain multiple loci, we will treat t as an indivisible unit since those loci are linked by inversions. Heterozygous males pass t to their offspring with frequency $\tau \geq 0.5$. Heterozygous females transmit t and + to their offspring with equal probability. This model has two variants. In the first variant, t/t males are sterile and t/t females have no reduction in fertility. In the second variant, t confers recessive lethality; no t/t animals of either sex are found in the population.

The most well studied autosomal USGC is *SD* in *Drosophila melanogaster* [23, 24, 22, 42, 47, 57]. Like the sterile variant of t, *SD* is characterized by transmission ratio distortion and infertility in homozygous males. The sterile variant of this model applies equally well to *SD* or to any other USGC with those two characteristics. Unlike mice, fruitfly populations are large and unstructured. On the population genetic level, the major difference between t and *SD* is population size and structure. Our model allows us to predict how differences in effective population size result in different evolutionary pressures on t and *SD*. The observed properties of these species in the wild can then provide supporting or contradictory evidence for these predictions.

3 A Markov Model of Genetic Drift in Populations of USGC's

In this section, we develop a theoretical framework for reasoning about genetic drift in populations of USGC's. We consider the simplest population model: genetic drift in an unstructured population of N individuals. It can model structured populations by replacing N with the effective population size. In addition, the model is useful for interpreting more complex modeling efforts. Simulation studies have modeled such populations as small populations linked by rare migration events [14, 28, 40]. In order to fully understand the behavior of these complex models, it is necessary to understand the role of genetic drift in the small populations that are their basic components.

This framework is based on the Wright-Fisher model, a Markov model of genetic drift in populations of non-distorting alleles which we briefly review below. Next we define analogous Markov models of both lethal and sterile distorters. An understanding of the state structure and the transition matrices of the models provides a great deal of intuition concerning the evolution of a finite population of distorters over time. Unfortunately, these models do not lend themselves to the standard approximation techniques for obtaining analytical estimates of the time to fixation. In the following section, we present the results of estimates of time to fixation obtained using Monte Carlo simulation.

3.1 Classical models of genetic drift: The Wright-Fisher Model

The Wright-Fisher model assumes a finite population of N diploid, monoecious individuals, resulting in $M = 2N$ alleles at any given locus. A single locus with two possible alleles (A and a) is considered. The population is assumed to have an initial A allele frequency, p_0.

The variation in allele frequency over time can be described by a Markov chain with $M+1$ states, where state j refers to a population with j A's and $M-j$ a's. Two states are absorbing ($j = 0$ and $j = M$). In other words, the chain has two possible equilibria: fixation of a and fixation of A. The transition probability of this Markov chain is:

$$P_{jk} = P(k|j) = \binom{M}{k} (\frac{j}{M})^k \left(\frac{M-j}{M}\right)^{M-k}. \tag{1}$$

The expected time to fixation can be estimated using the diffusion approximation (see, for example, [16]) and depends linearly on M, the number of alleles in the population:

$$E[t_f] = 2M \cdot (p_0 \log p_0 + (1-p_0) \log (1-p_0)).$$

The implications of this analysis are that small changes in population size lead to small changes in the expected fixation time. Allele frequencies in small populations vary widely from generation to generation and are not characterized by polymorphic equilibria.

3.2 Sterile distorters in finite populations

We can construct a similar Markov chain for a population of sterile distorters. As above, we assume a population of N diploid organisms. In this case, we focus on a distorting locus with two alleles (+ and t). Heterozygous (+/t) males exhibit a distorted transmission ratio, τ, and t/t males are unconditionally sterile. The t-allele has no unusual effects in females.

Variation over time in the frequency of the t allele can be modeled using a Markov chain generalizing the Wright-Fisher model. However, it is necessary to keep track of genotype frequencies, rather than allele frequencies because t has a different phenotype in t/t males (sterility) than in +/t males (TRD). This means that the state of the Markov chain in generation i is described by a four-dimensional vector $\{Y_m(i), Z_m(i), Y_f(i), Z_f(i)\}$, where

$$\begin{aligned}
Y_m(i) &= \text{the number of +t males in generation } i, \\
Z_m(i) &= \text{the number of tt males in generation } i, \\
Y_f(i) &= \text{the number of +t females in generation } i, \\
Z_f(i) &= \text{the number of tt females in generation } i.
\end{aligned}$$

We assume that the number of males in each generation is fixed at N_m and the number of females at N_f, yielding a total of $N = N_m + N_f$ individuals at each generation. This implies that, in generation i, the number of +/+ males is $X_m(i) = N_m - Y_m(i) - Z_m(i)$ and the number of +/+ females is $X_f(i) = N_f - Y_f(i) - Z_f(i)$. For notational convenience, we define the number of +/+ individuals of either sex in generation i to be $X(i) = X_m(i) + X_f(i)$, with similar definitions for $Y(i)$ and $Z(i)$.

The population will either eventually consist entirely of +/+ individuals (both male and female) or will become extinct (because males in some generation are all t/t). We wish to consider the probabilities of these two outcomes and the mean time until one outcome or the other occurs.

The probabilities q_m (q_f) that a gene transmitted from a male (female) of generation i is a t allele are

$$\begin{aligned}
q_f &= \frac{Y_f(i)}{2N_f} + \frac{Z_f(i)}{N_f}, \\
q_m &= \frac{Y_m(i) \cdot \tau}{X_m(i) + Y_m(i)}.
\end{aligned}$$

Thus the respective probabilities $P(++)$, $P(+t)$ and $P(tt)$ that an embryo (male or female) in the next generation is +/+, +/t and t/t are given by

$$\begin{aligned}
P(++) &= (1 - q_m) \cdot (1 - q_f), \\
P(+t) &= (1 - q_m) \cdot q_f + (1 - q_f) \cdot q_m, \\
P(tt) &= q_m \cdot q_f.
\end{aligned}$$

This leads to the following transition probability for the Markov chain

$$P(Y_m(i+1), Z_m(i+1), Y_f(i+1), Z_f(i+1) \mid Y_m(i), Z_m(i), Y_f(i), Z_f(i)) = \frac{N_m! N_f!}{X_m(i+1)! Y_m(i+1)! Z_m(i+1)! X_f(i+1)! Y_f(i+1)! Z_f(i+1)!} P(++)^{X(i+1)} P(+t)^{Y(i+1)} P(tt)^{Z(i+1)}$$

Because four state variables are required to describe every state, the one dimensional diffusion approximation cannot be used to obtain a tractable, analytical expression for the expected time to loss of t. Instead we used Monte Carlo simulation, described in Section 4, to obtain estimates for the expected time to absorption.

3.3 Lethal distorters in finite populations

A Markov model of genetic drift of lethal USGC's, where both male and female *t/t* embryos die before birth, can also be developed. The consequence of lethality is that the highest t-haplotype frequency is $q = 0.5$ and occurs in the state where all individuals are heterozygotes. This results in a Markov chain with $N+1$ states, with only one absorbing state (where all alleles are $+$).

The probability of having k heterozygotes in generation $i+1$, given j heterozygotes in generation i can be shown to be [13]

$$P_{jk} = P(k|j) = \binom{N}{k} (Q'(j))^k \, (1-Q'(j))^{N-k},$$

where

$$Q'(j) = \frac{-8\, q(j)^2\, \tau + 2\, q(j)\ (1+2\,\tau)}{2-4q(j)^2\, \tau}$$

is the frequency of heterozygotes after selection. Q' is derived from $q(j)$, the t-haplotype frequency in generation i, by observing that the probability of obtaining a heterozygote is the probability that an embryo will be heterozygous normalized by the probability that an embryo will survive (i.e., is not *t/t*).

We could now apply the one dimensional diffusion approximation by obtaining an expression for the expected mean and variance of P_{jk}. Instead, a simpler approximation was developed. By modeling the progression of time in terms of birth/death events instead of generations, the expected time to fixation can be estimated based on a continuant model analogous to the Moran haploid model [16]. This yields an open form expression for the time to fixation which can be evaluated numerically to show the dependence of the mean fixation time on the transmission distortion and the population size. The results of this model are presented in a companion paper [13]. In the current paper, we present Monte Carlo simulations of the progression of lethal allele frequencies as a function of generation for comparison with the results for steriles.

4 Simulation Results

In the previous section, we made qualitative observations concerning genetic drift in populations of USGC's by constructing Markov models analogous to the neutral Wright-Fisher model of non-distorting alleles. The Wright-Fisher model tells us that there are two possible outcomes for a population of non-distorting alleles: loss of a and loss of A. In the absence of selection, a and A are equally likely to prevail given an initial allele frequency of 50% and these two outcomes are symmetric. In the sterile model, there are also two possible outcomes: wild type fixation and extinction. However, these two outcomes are not symmetric in that the fixation of t results in a scenario which is disastrous for the population and for t itself. For lethal distorters, only one outcome is possible: wild type fixation. The t-allele frequency can never rise above 50%. In other words, recessive lethality effectively chops off one end of the Markov chain.

All equilibrium conditions in both the sterile and lethal models result in loss of t, either through wild type fixation or through extinction. Thus the expected time to reach a steady state is a measure of the intensity of negative selective pressure on t due to sterility or lethality: the longer it takes to reach an equilibrium state, the longer t can survive. In the Wright-Fisher model, an equilibrium condition is reached fairly quickly: the time to absorption grows linearly with population size. How quickly will equilibrium be reached in populations of USGC's?

We address this question using Monte Carlo simulation. Our simulator models a population of N individuals, each represented by sex and genotype (+/+, +/t and t/t, in the case of sterile alleles). The population size, N, is fixed from generation to generation and the sex ratio is fixed at 50:50. Each generation is calculated from the previous generation, one individual at a time. For each offspring, two parents are selected at random and the offspring genotype is calculated from the parental genotypes following the genetics of transmission ratio distortion. If the offspring is not viable (because the father is sterile or because it carries two lethal t's) the program tries again. The simulation of new generations is repeated until t is lost, either through wild type fixation or through extinction. The variation of t allele frequency as a function of generation and the time to loss of t were measured as a function of τ and N. In the case of steriles, the manner in which t was lost was also recorded. Preliminary experiments indicated that initial t-allele frequency has very little impact on the final outcome, so an initial allele frequency of 50% was always used.

Expected time to fixation as a function of TRD and population size is shown in Figures 1 – 3 for the sterile model and Figures 6 and 7 for the lethal model. Each point is the average of 20 runs. Error bars in the figures represent 90% confidence intervals. Figure 1 shows the impact of τ on expected time to loss of t when $N = 30$. For low and high values of τ, the population reaches an extreme state within tens of generations. However, for intermediate values of τ ($0.70 < \tau < 0.85$) the expected time to loss of t rises quickly to thousands of generations.

Figures 2 and 3 show that this effect becomes more pronounced as population size increases. For populations of size ten, fixation occurs quickly for all values of τ, which is

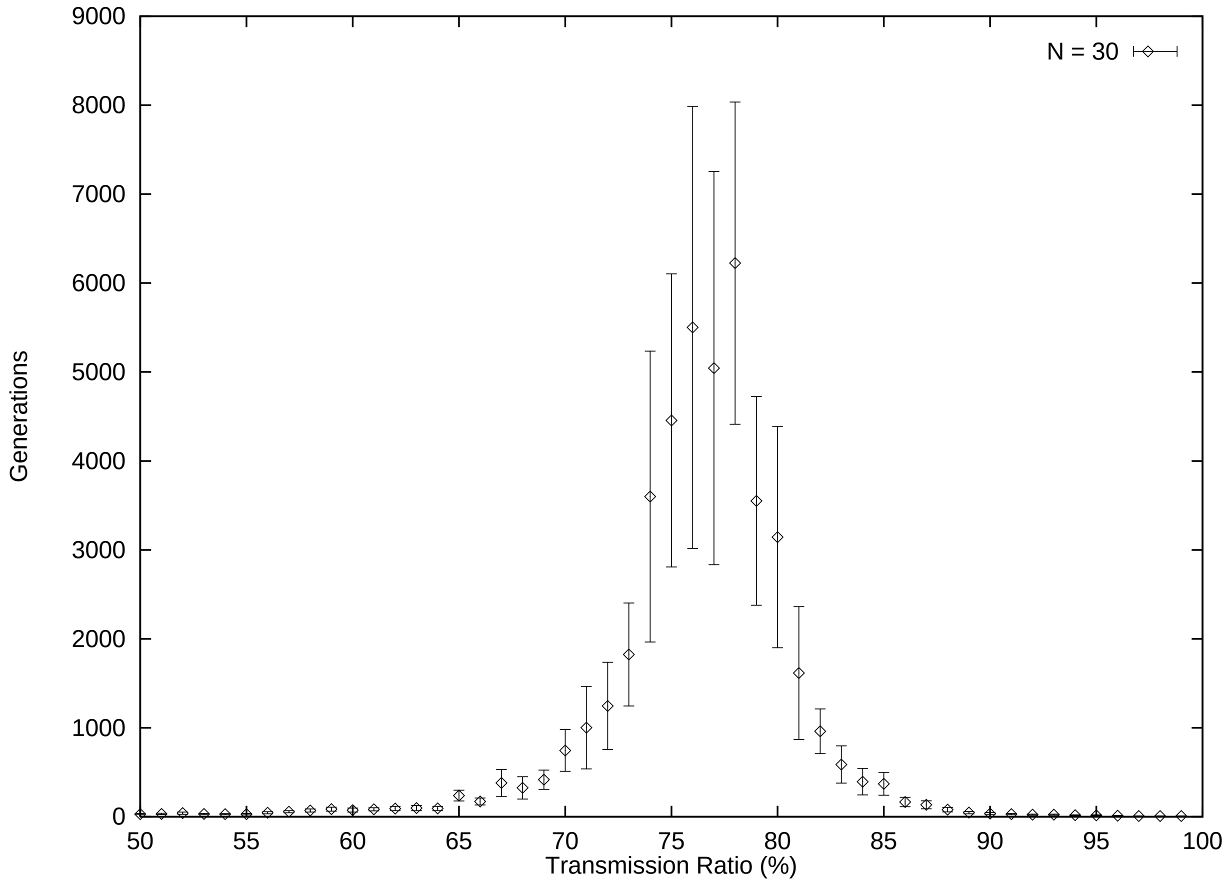

Figure 1: Expected time to loss of sterile USGC's in a population of 30 individuals as a function of transmission ratio. Each point is averaged over twenty experiments. Error bars represent 90% confidence intervals.

consistent with the observations of Lewontin [29]. However, the time to fixation grows exponentially as N increases, resulting in a peaked bell curve. As population size increases, this peak also moves slowly to the right.

The short absorption times at low TRD occur when the system is rapidly captured by the absorbing state corresponding to wild type fixation. Short absorption times associated with high TRD occur when t quickly spreads through the population, leading to extinction. This can be seen in Figure 4, which shows the fraction of experiments which terminated through wild type fixation or extinction, respectively. For low TRD, wild type fixation occurs in most experiments. For high TRD, extinction is the prevalent outcome. A transition zone where both outcomes are possible is seen in the middle. This transition moves to the right and becomes more abrupt as the population size increases. Notice that for each value of N, the crossover point is located at roughly the same value of τ as the peak of the corresponding curve in Figure 2.

This behavior can be understood by looking at Figure 5, which shows t-allele frequency as

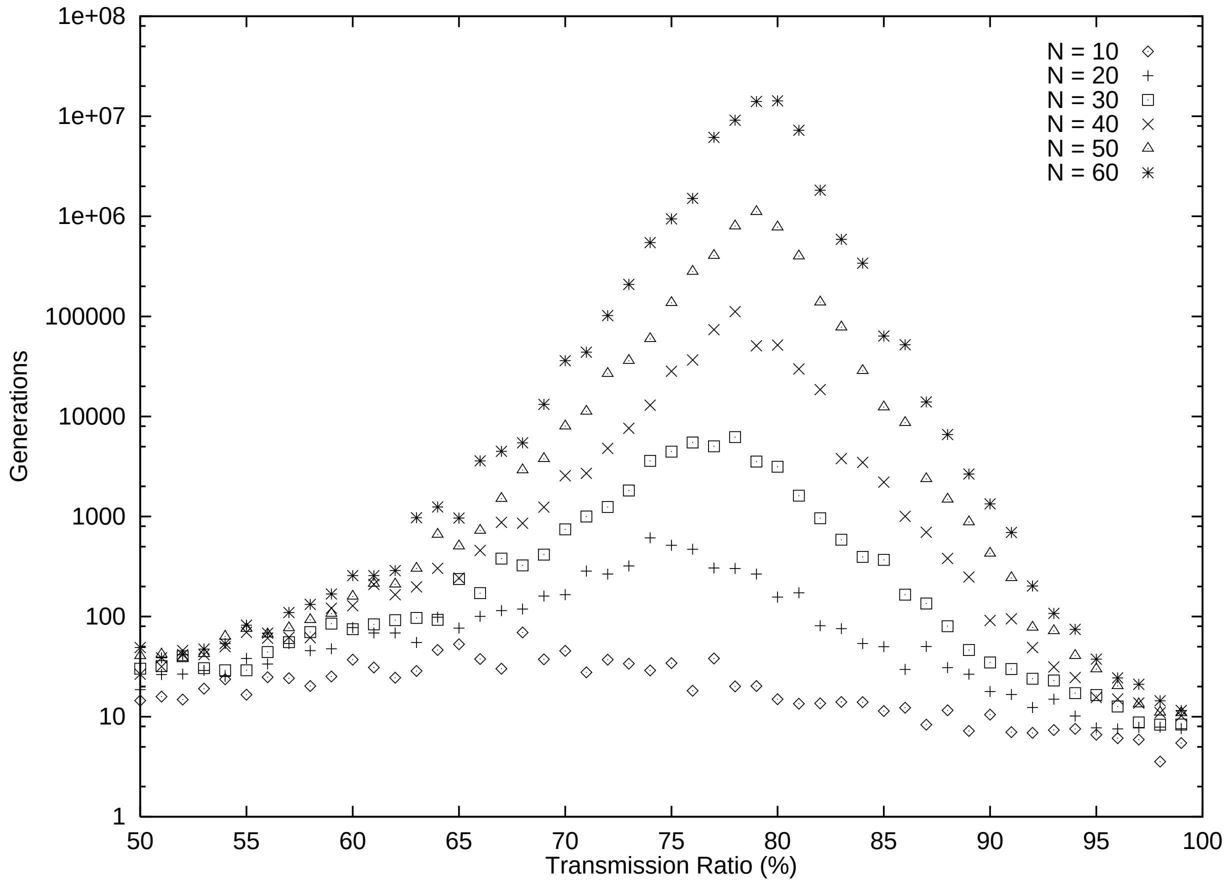

Figure 2: Expected time to loss of sterile USGC's as a function of τ for varying population sizes. Each point is the average over twenty experiments.

a function of generation for $\tau = 0.70$, $\tau = 0.78$ and $\tau = 0.85$ in a population of size 40. Each point is the average allele frequency, over all experiments, seen at the current generation. The bars represent the minimum and maximum value of q seen at that generation in any experiment. Taken as a whole, each curve shows the degree of variation seen as the simulation progresses.

The values $\tau = 0.70$, $\tau = 0.78$ and $\tau = 0.85$ were chosen because they correspond approximately to the lower limit, peak and upper limit of the bell in the curve in Figure 2. Figure 5 shows that the average allele frequency increases with τ. For the lower limit of the bell curve ($\tau = 0.70$), the average allele frequency is low enough so that the bottom of the range is close to $q = 0$, leading to wild type fixation. Similarly, for the upper limit of the bell curve ($\tau = 0.85$), the top of the range approaches $q = 1.0$, which would lead to extinction. However, for $\tau = 0.78$, near the center of the peak, the chance of reaching either absorbing state is very unlikely.

Figures 6 and 7 show the impact of N and τ, respectively, on expected time to wild type fixation in populations of lethal t-haplotypes. For lethals, the time to fixation grows

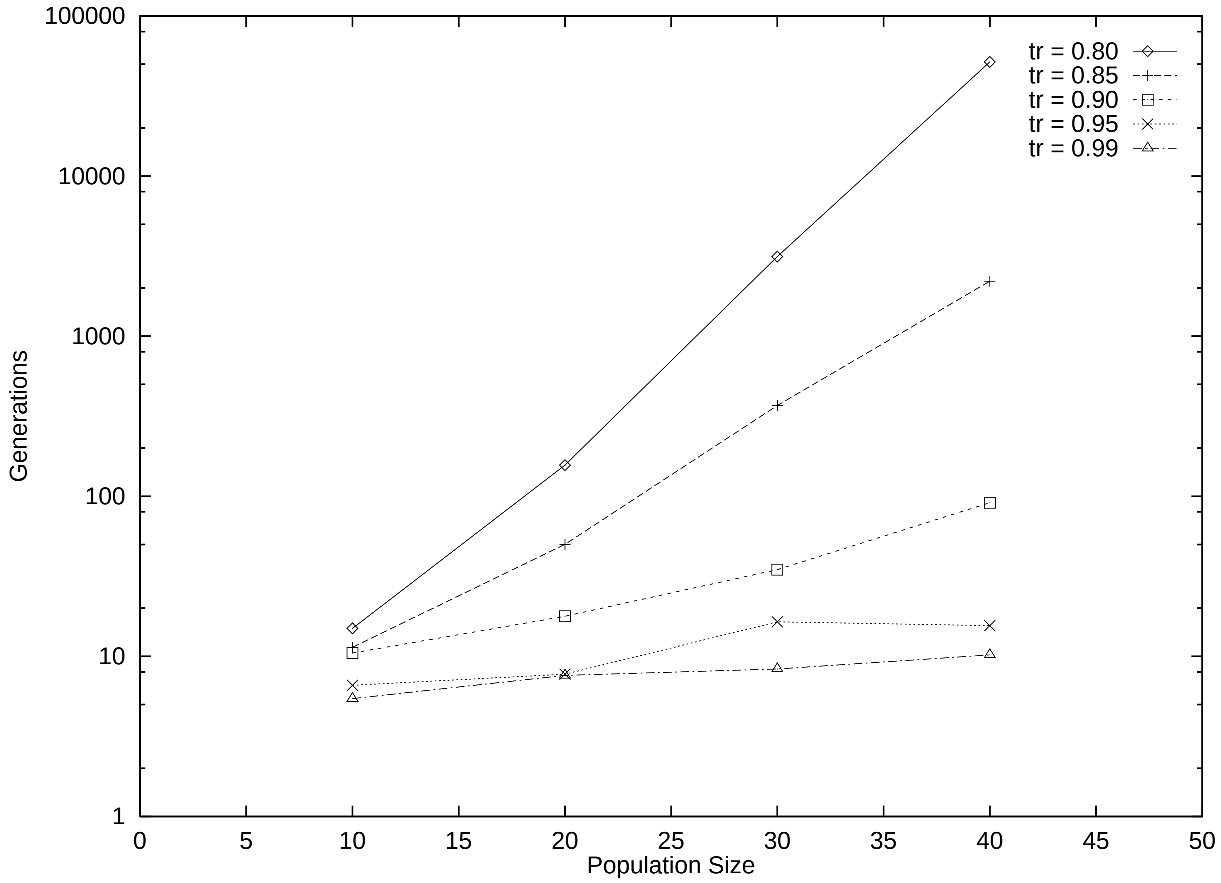

Figure 3: Expected time to loss of sterile USGC's as a function of population size for varying transmission ratios. Each point is the average over twenty experiments.

exponentially with both N and τ. This is due to the absence of a second absorbing state at high t-allele frequencies. The lethal case differs from the sterile case in that the greater the transmission distortion, the longer the USGC survives.

Taken together, the simulations results of the sterile and lethal models show important differences between genetic drift in populations of non-distorting alleles and USGC's. In the non-distorting (Mendelian) case, the expected time to fixation depends linearly on population size as demonstrated by the neutral Wright-Fisher model stated in Equation 1 and the allele frequency varies substantially until fixations occurs. In contrast, the expected time to loss of t varies non-linearly with τ and N, so that a small change in population size or transmission ratio can result in a dramatic change in the dynamics of the system. Furthermore, populations of USGC's exhibit quasi-stable states. We use the term "quasi-stable" to describe a state in which the time to absorption is on a different time scale than that for the standard Wright-Fisher selectively neutral case. This difference is larger by at least one order of magnitude, resulting in a long-term polymorphism whose allele frequency is characterized by a well-defined mean and variance (data not shown). This quasi-stable state

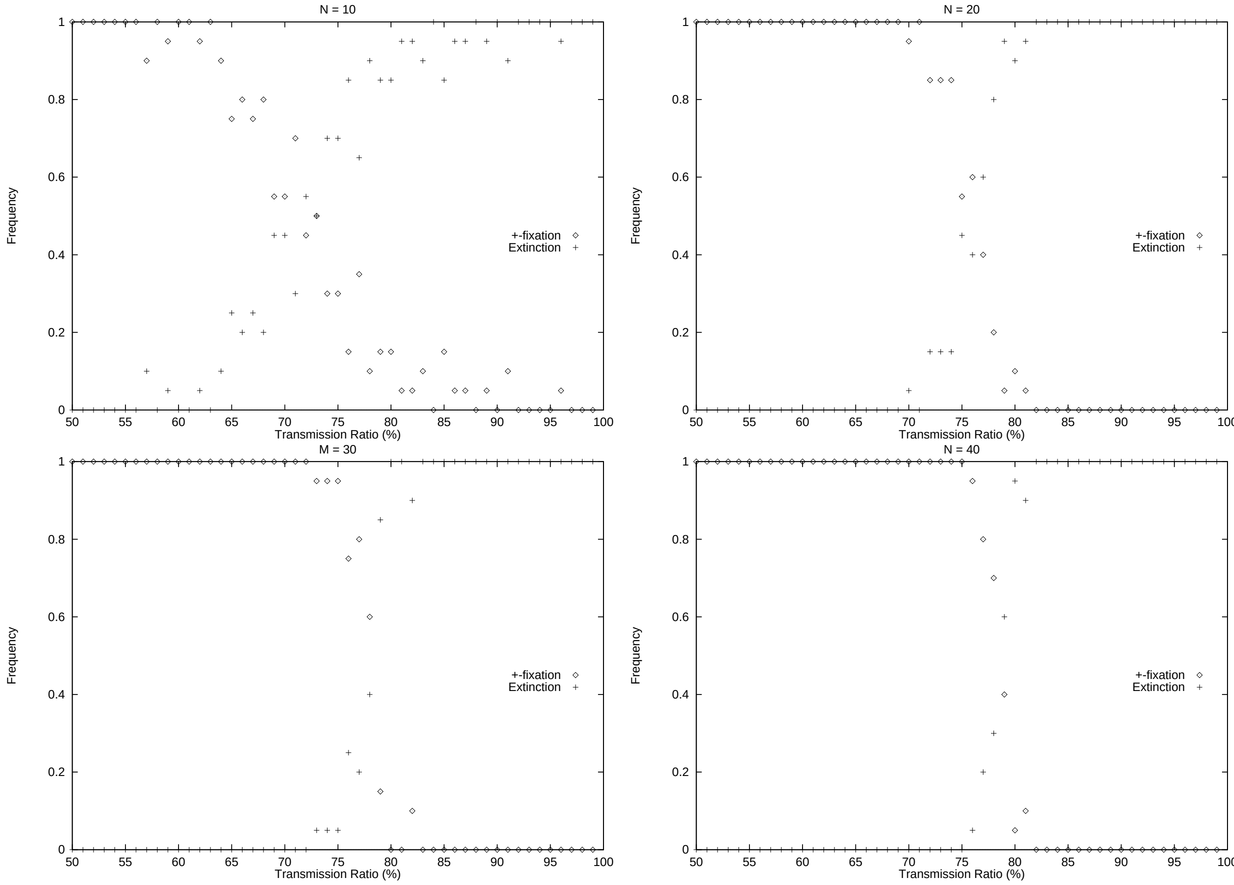

Figure 4: Frequency of expected outcome (wild type fixation or extinction due to sterility) of fixation experiments as a function of transmission ratio for populations of 10, 20, 30 and 40 individuals, respectively.

occurs at intermediate values of τ for steriles and at high values of τ for lethals.

Computer simulations of finite populations are by their nature stochastic. This leads to questions such as: (1) when is the simulator in a quasi-stable state? (2) when should the allele frequency be measured? and (3) how long should the simulation be allowed to run? Our results delineate ranges of N and τ for which distorting alleles will persist (i.e., those parameter values for which a quasi-stable state occurs) and parameter values for which the system is unstable, (i.e., regions with non-linear behavior where small changes in N or τ can result in a large change in allele frequency). Our results also suggest when and how to measure allele frequency. In regions of the parameter space where a quasi-stable state occurs, allele frequency can be characterized by the mean and variance. Outside those regions, the allele frequency is not stable. Instead, we measure whether t fixes or is lost from the population and the mean time until this event occurs.

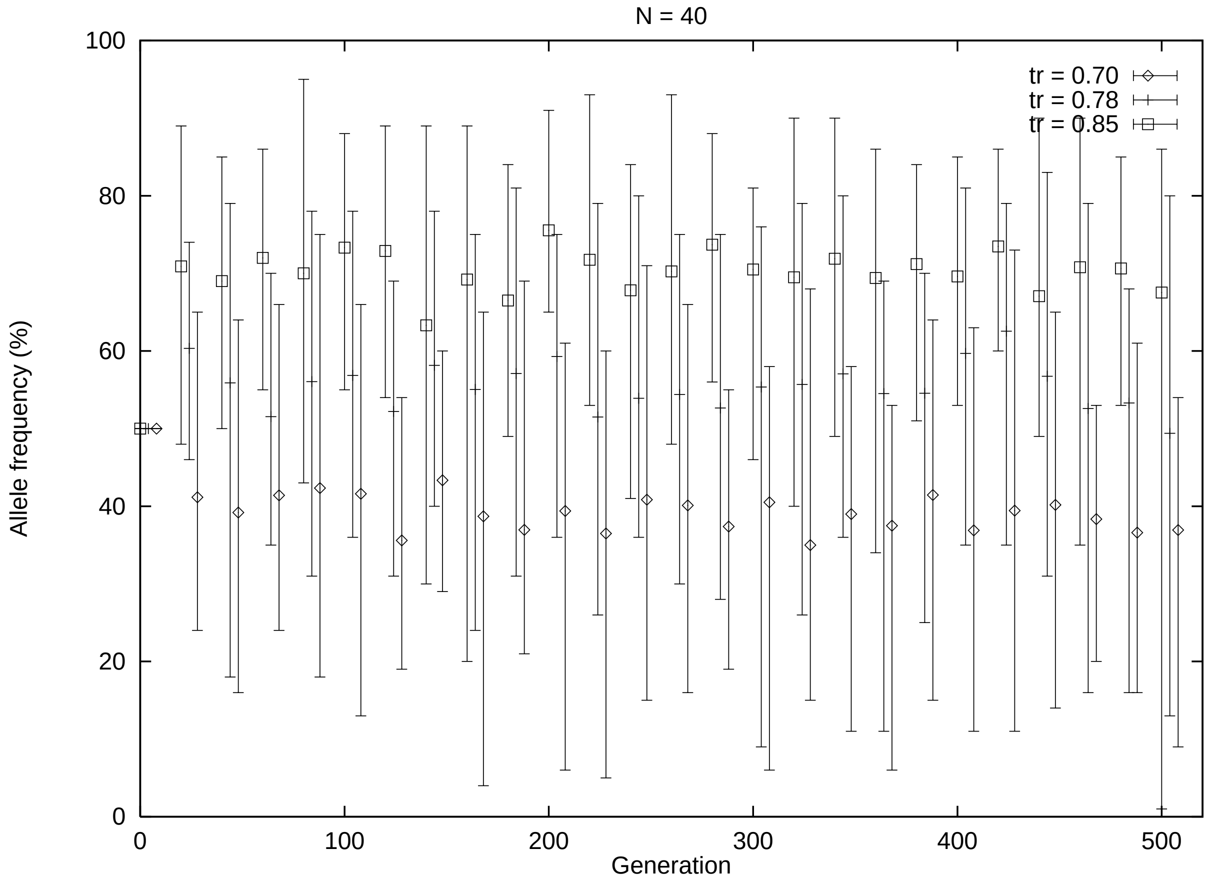

Figure 5: Variation in average frequency of sterile USGC's over time. Each point represents the ultraselfish allele frequency seen at the current generation averaged over twenty experiments. The bars represent the minimum and maximum allele frequency seen at the current generation in any of the twenty experiments.

5 Related Work

Most previous models of the t-haplotype have focussed on questions of population dynamics: what forces in modern day mouse populations account for the t-haplotype frequency seen in nature? The earliest population models assumed random mating in infinite populations. Based on these assumptions, Bruck [6] predicted allele frequency as a function of transmission ratio distortion in populations of lethal t-haplotypes:

$$q = .5 - \sqrt{\tau(1-\tau)}/2\tau. \tag{2}$$

This panmictic model predicts a t-haplotype frequency of 0.333 when $\tau = 0.9$, which is much higher than the empirical range of 0.1 - 0.15 [2, 4, 26, 46]. A similar model for sterile t-haplotypes

$$q = 2\tau - 1 \tag{3}$$

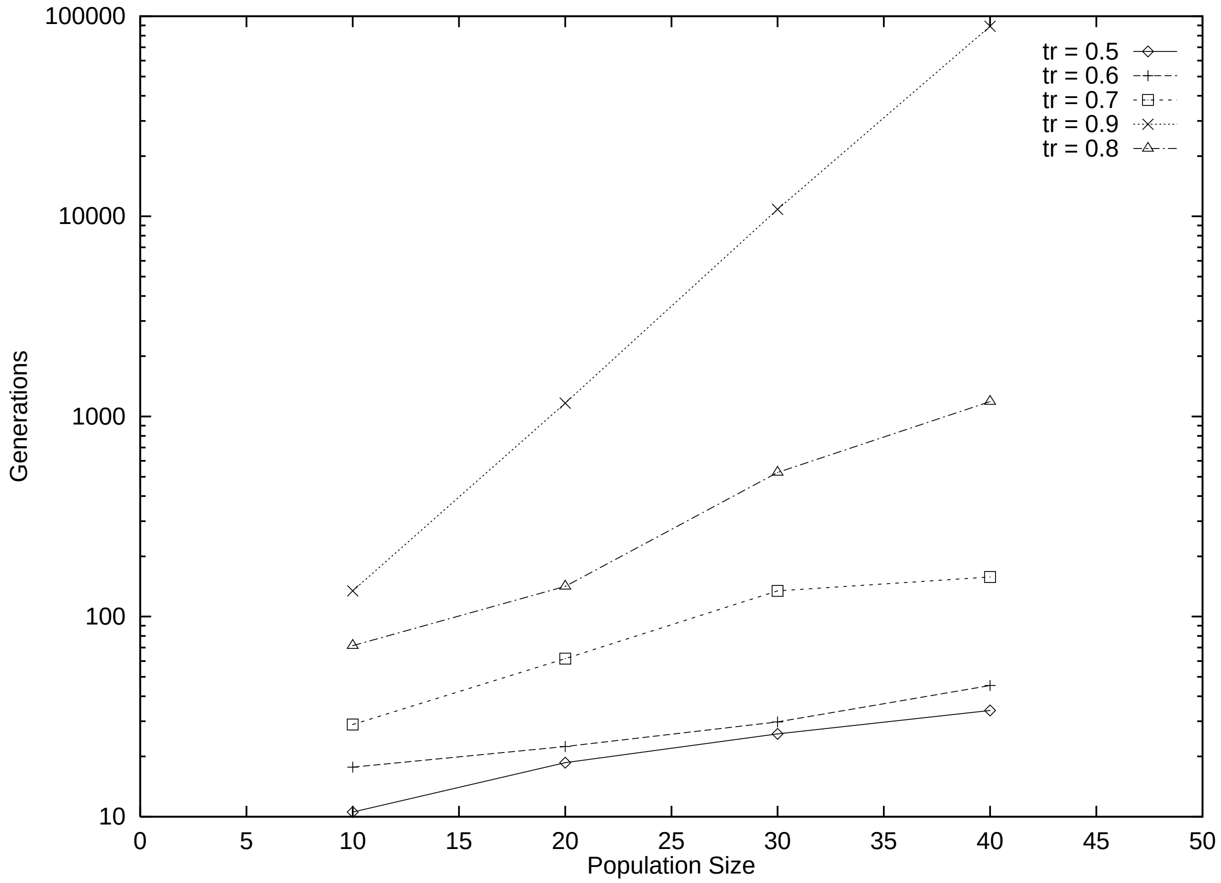

Figure 6: Expected time to loss of lethal USGC's as a function of population size, for varying transmission ratios. Each point is the average over twenty experiments. Error bars represent 90% confidence intervals.

was presented by Dunn and Levene [12]. This model also predicts t-haplotype frequencies that are much greater than those seen in the wild, suggesting that infinite population models are not good approximations for mouse populations.

In the early sixties, Lewontin and Dunn [30, 29] studied the behavior of t in finite populations using Monte Carlo simulation. They did not perform a comprehensive study of the stochastic behavior of t-haplotype frequencies as a function of transmission ratio and population size but the cases that they studied provided good intuition concerning the population dynamics of t and was an inspiration for the current work. In particular, they observed that for small demes (approximately ten mice), wild type fixation occurs relatively rapidly whereas for larger demes ($20-50$ mice) lethal t haplotypes will persist over many generations. From this observation, they hypothesized that the low level t polymorphism observed in nature is due to a balance between the loss of t in small demes due to genetic drift and the reintroduction of t through interdemic migration.

This hypothesis was widely accepted for thirty years. In separate studies, Levin *et al.* [28]

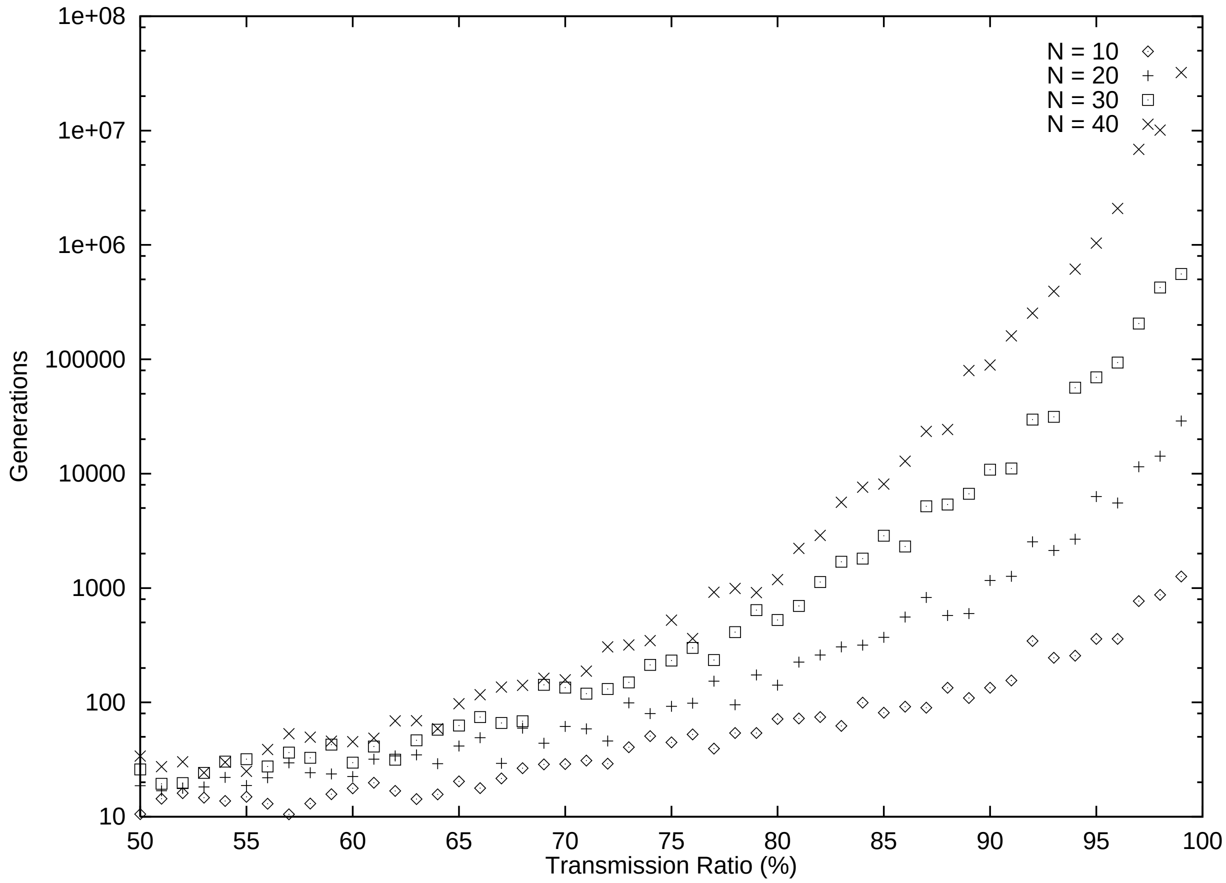

Figure 7: Expected time to loss of lethal USGC's as a function of transmission ratio for varying population sizes. Each point is the average over twenty experiments.

and Nunney and Baker [40] presented Monte Carlo simulations of structured populations in which they demonstrated that a deme size and mutation rate could be found that, taken together, predict the t-haplotype frequency seen in nature. However, in 1997, Durand *et al.* [14] showed that although it is possible to find such values, they are not stable with respect to small perturbations in the interdemic migration rate. If the interdemic migration rate is decreased slightly, t will be lost from the population. If it is raised slightly, the allele frequency will rapidly increase to the high frequencies predicted by the panmictic model. This instability in allele frequencies is not consistent with natural observations and suggests that a balance between drift and migration is not a sufficiently powerful model to explain the stable t-polymorphism seen in nature.

While the focus of the current work is on evolutionary questions, the results reported in the previous section have some relevance to the population dynamics work surveyed here. They show that genetic drift is fundamentally different in population of USGC's in that expected absorption time varies nonlinearly with N and τ and in that, for certain regions of the parameter space, a quasi-stable state can exist. This explains Lewontin's observation that t disappears in small populations but persists in larger ones. Our results also suggest why a balance between the loss of t-alleles due to genetic drift and reintroduction of t-alleles through interdemic migration does not result in a *stable*, low-level t-haplotype polymorphism. A change in interdemic migration rate is analogous to a change in effective population size. The results in Figure 6 show that small changes in population size can result in large changes in persistence time. Thus, a small change in migration rate could move the population from a regime where the t allele is lost in 10 to 100 generations to a quasi-stable state where it persists for a long time.

In addition to work on the population genetics of t-haplotypes, there is a substantial body of theoretical analyses of ultraselfish genes in other organisms (see [18] for a survey.) This, which deals entirely with infinite populations, includes one-locus models of the sex-ratio (SR) chromosome in *Drosophila pseudoobscura* [15, 9, 52] and the Segregation Distorter (*SD*) complex in *Drosophila monogaster* [23] and two-locus models of *SD* [43, 8, 20, 31, 51] and SR distortion in *Aedes aegypti* [37, 27, 17]. The role of recombination and selection in modifiers of SR has also been studied [27, 17]. An analysis of the evolutionary dynamics of spore killer genes has been presented recently [39].

Two papers have been published that examine the evolution of lethal t-haplotypes by considering a population in which two or more t-haplotypes compete. Charlesworth [7] has studied the allele frequencies of a wild type, a sterile and one or more lethal t-alleles in an infinite, panmictic population model involving reproductive compensation and kin selection. van Boven *et al.* [55] examined a system involving complementation between alleles with different degrees of distortion.

There has been very little work on analytical models of USGC's in finite populations. van Boven [54] used a Markov analysis to compare the persistence of USGC's to that of non-distorting alleles using Robertson's concept of a retardation factor [45]. This factor is the ratio of the largest non-unit eigenvalue in the ultraselfish case to the corresponding eigenvalue of the neutral Wright-Fisher model (Equation 1). A disadvantage of this approach is that it only gives a good approximation if the other non-unit eigenvalues are much smaller and hence higher powers of these eigenvalues approach zero rapidly as time increases. van Boven has not shown this to be the case. Whereas his predictions for sterile alleles are similar to ours, his predictions for lethal t-haplotypes differ. Unlike the results shown in Figures 6 and 7, van Boven's predictions of time to fixation for lethals do not show exponential increase as τ increases. This is at odds with the results presented in the current paper and in [13], which were obtained using two different approaches. Intuitively, we would expect an increase in mean absorption time with increasing τ. This is because the Markov chain describing populations of lethal t-haplotypes has only one absorbing state, corresponding to wild type fixation, and the probability of reaching this state drops dramatically as τ increases. This discrepancy between our predictions and those of van Boven's approach could be due to the limitations of using only the largest non-unit eigenvalue to analyze the convergence of USGC frequencies.

6 Discussion

The results in Section 4 show that, for both the lethal and sterile models, there are parameter values for which a quasi-stable state exists. This quasi-stable state occurs at intermediate values of τ for steriles and at high values of τ for lethals. In both cases, distorting alleles with values of τ that place them in the quasi-stable region have a selective advantage over those that are rapidly lost. What does this imply for the evolution of distorting alleles?

Below we argue that genetic drift plays an important role in the evolution of distorting alleles, resulting in different evolutionary pressures in panmictic species and species with small and/or structured populations. We present a hypothesis, based on our results, to explain why the transmission ratio in the t-haplotype observed in the wild is 0.9 and why recessive lethal mutations are a recent but highly successful innovation. The same arguments can also explain why $\tau \approx 0.99$ in Drosophila and why lethal mutations are rarely associated with *SD*.

6.1 Evolution of Transmission Ratio Distortion

It is likely that the transmission ratio in early t-haplotypes was just above 0.5 and increased slowly over time as additional inversions containing alleles that contribute to distortion were recruited to the t-haplotype and as additional mutations in the initial distorting alleles increased their effectiveness. Since lethals observed today arose only recently, we assume TRD originally evolved in populations of sterile t-haplotypes. Under this hypothesis, the transmission ratio observed in modern day t-haplotypes is the result of the balance between the positive and negative selective forces on TRD. Positive selection promotes increased TRD since variants with a greater degree of distortion will be transmitted to more offspring. Genetic drift combined with sterility exerts a negative selective force, since for distorting alleles, drift always results in the loss of t. What is the exact nature of this negative selective force?

Variants of t within the quasi-stable region in the curves in Figure 2 clearly have an advantage over those outside the region. In particular, sterile t alleles will not evolve values of τ corresponding to the region where the expected absorption time is on the order of 10 to 100 generations. Which variants within the quasi-stable region will be most successful? Theoretically, those alleles with the longest persistence time (i.e., variants whose value of τ corresponds to the maximum of the peak) will be most successful. However, in the real world, negative selective pressure on transmission ratio will not be sensitive to variations in persistence times in the range of millions of years. Rather, negative selective pressure is uniformly negligible in the quasi-stable region, depicted in the cartoon in Figure 8. For one thing, for populations of sixty or more individuals, the persistence time within the quasi-stable region is long compared with two to three million years, the phylogenetic estimate of the age of the t-haplotype [19, 38, 50]. It is also long when compared with the rate of change in transmission ratio due to evolutionary forces. Furthermore, genetic drift will only influence TRD evolution during periods when the population size and structure are stable.

Within the quasi-stable region, catastrophic events, such as famine or migration, are more likely to eradicate a population than extinction due to male sterility, masking any impact variations in persistence time might have on the evolution of distortion. Therefore, the location of the boundaries of the quasi-stable region, and not the shape of the peak, is the primary factor in the evolution of TRD.

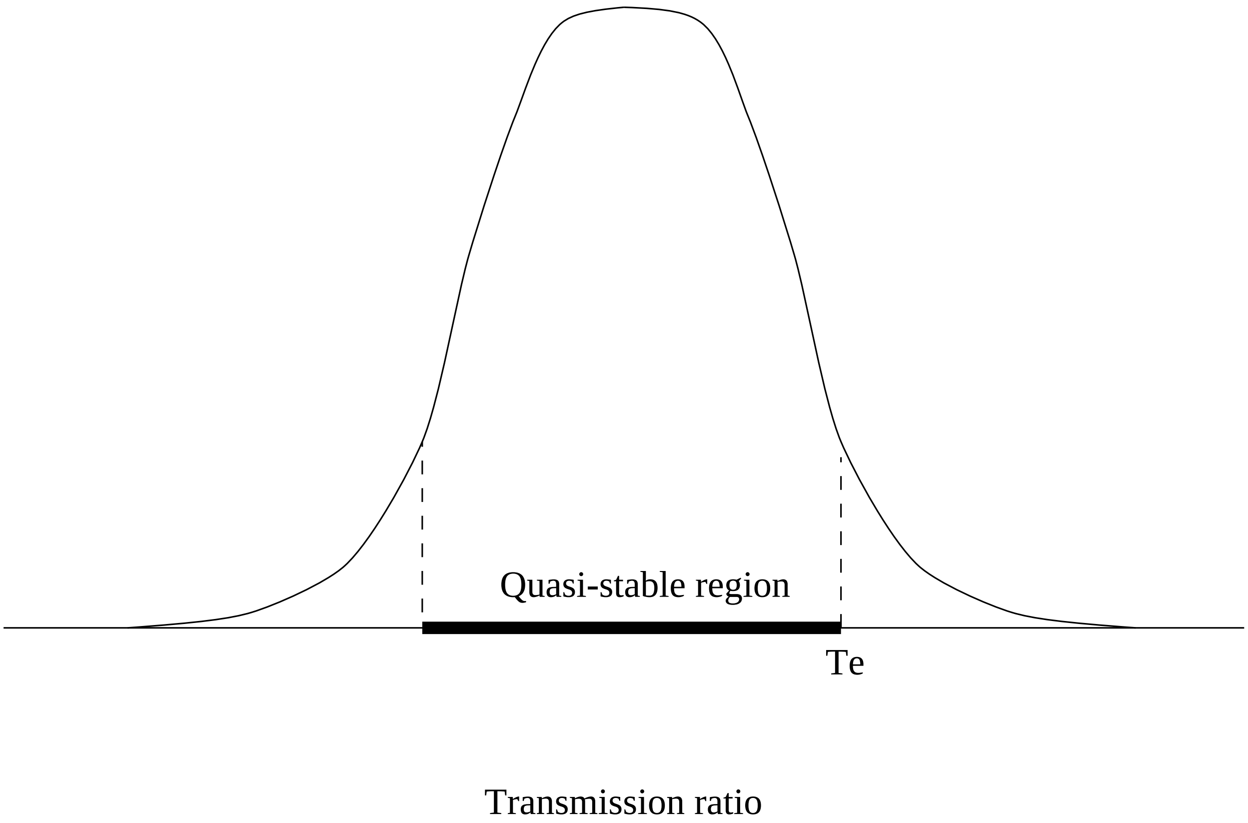

Figure 8: An abstraction of persistence time of a sterile USGC as a function of transmission ratio. Negative selective forces on transmission ratio within the interval, W, are negligible. Postive selective forces assert pressure to increase the transmission ratio, while the abrupt reduction in persistence time on the right hand side of W exerts a negative selective force on the transmssion ratio. Thus transmission ratio will increase until it reaches the value, T_e, on the right extreme of W.

Thus according to our hypothesis, TRD will increase under positive selective pressure until it reaches T_e, the transmission ratio at the right bounadary of the quasi-stable region, as shown in Figure 8. We used our simulator to determine the location of this boundary as a function of population size (Figure 9), yielding $T_e = 0.87$ for $N = 50$, $T_e = 0.93$ for $N = 100$ and $T_e > 0.99$ for $N = 500$. Figure 9 shows that although in species with effective population sizes above 500, such as Drosophila, male infertility does not exert a negative selective force on the transmission ratio, in species with effective population sizes of 100 or less, such as mice, homozygous male sterility will limit the evolution of the transmission

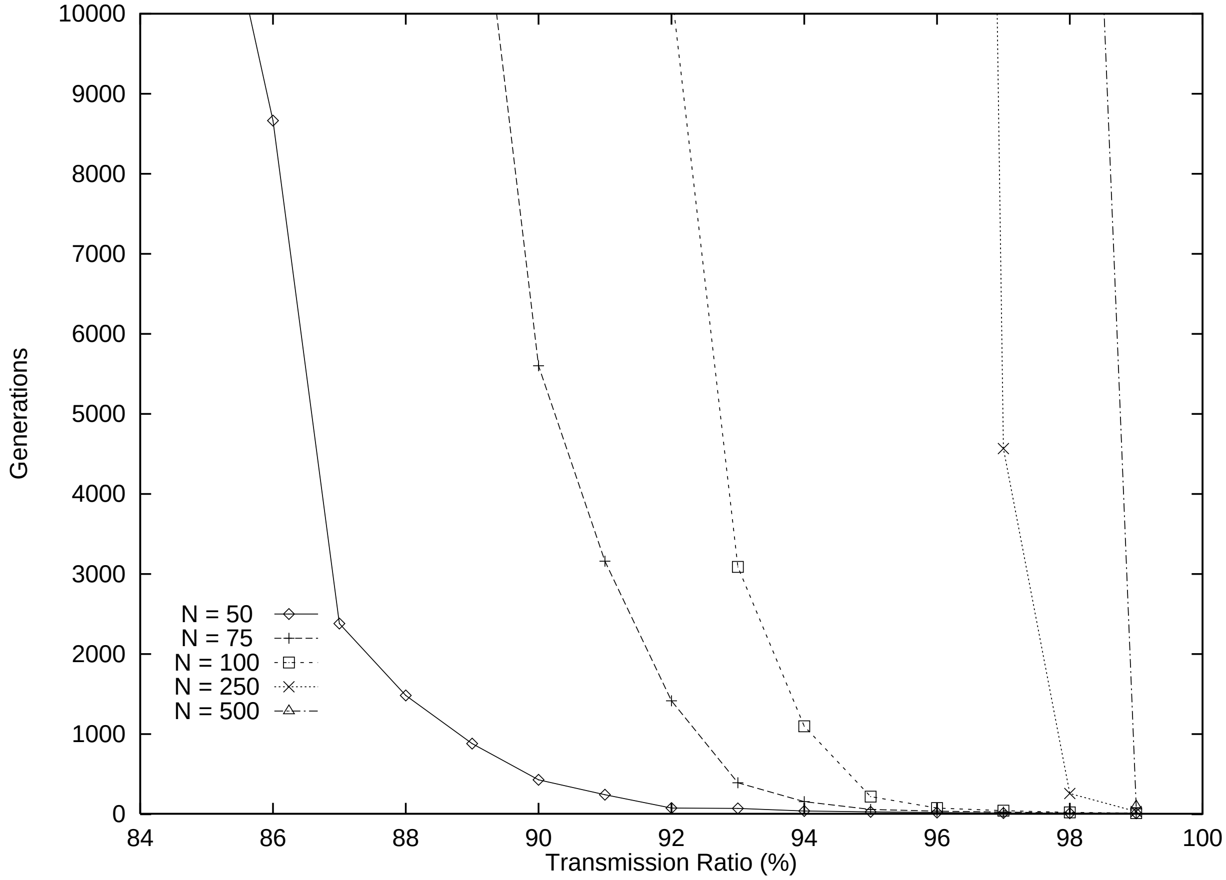

Figure 9: Expected time to loss of sterile USGC's for large populations in the high τ region. Populations of 500 or more individuals see no negative selective pressure due to male sterility. In populations of less than 100 individuals, a value of $\tau \geq 0.95$ will lead to extinction.

ratio. This is consistent with the mean observed modern day transmission ratio of $\tau \approx 0.9$ in mice [3]. Figure 9 also explains why $\tau \approx 0.99$ in Drosophila since the effective population sizes for those species are much larger than 500. In this case, genetic drift does not exerts a negative selective force which limits TRD, allowing the transmission ratio to approach 1.0.

The presence of lethal mutations effectively eliminates the absorbing state associated with extinction. As seen in Figure 7, once lethals appear, persistence time no longer decreases with high TRD. Since most contemporary populations contain lethal t-haplotypes, why hasn't τ increased beyond a mean of 0.9? A possible explanation stems from the fact that the lethal mutations found on the t-haplotype have a simple basis, so that the transition from viability to lethality can be rapid. Transmission ratio distortion, which is due to complex interactions between genes, evolves much more slowly. It is possible that the transmission ratio is now free to increase due to the advent of lethal mutations, but 10,000 years is not long enough to see an appreciable increase. In addition, τ may have been lower 10,000 years ago when these lethals first appeared.

6.2 Appearance of lethal t haplotypes

Our results also suggest a hypothesis for the recent but widespread appearance of lethal, recessive mutations on the t-haplotype. This same hypothesis also explains why lethals mutations are not generally associated with *SD*. In early t-haplotype populations with low TRD, wild type fixation was probably the primary cause of loss of both lethal and sterile t-haplotypes. The frequency of t is higher in populations of steriles and t/t females are able to pass t on to their offspring. Thus, recessive lethality confers no selective advantage in this region and, in fact, reduces the t-haplotype frequency. There may have been lethal mutations linked to early t-haplotypes; but, in the absence of any selective advantage, there is no reason why they should have persisted. However, lethals have a selective advantage in the region of Figure 4 where extinction is the prevalent outcome. Once TRD increases past the crossover point, populations of sterile t-haplotypes will be eliminated by male sterility. Populations of lethals will persist, however, and individuals from those populations can recolonize sterile demes. This is consistent with the recent appearance of lethals and can explain why most t-haplotypes found in the wild carry recessive lethal mutations. This is also consistent with the many different lethal variants found in wild populations. Our hypothesis suggests that, because of their selective advantage in the extinction zone, once lethal mutations arise they will persist.

Figure 9 suggests that the crossover point approaches $\tau = 1.0$ as population size increases. For this reason, we believe that lethal mutations do not have a selective advantage in large insect populations such as Drosophila populations. This would explain why lethal mutations are rarely seen in *SD*.

Acknowledgements

We thank Almut Burchard, Max Mintz and Ken Steiglitz for helpful discussions and the Princeton University Computer Science Department for welcoming DD as a visiting researcher. DD was supported by NSF grants BIR-94-13215 A01 and DEB-9752945. WJE was supported by NIH grant GM21135. LMS was supported by NIH grant HD20275.

References

[1] P.K. Anderson. Lethal alleles in *mus musculus*: Local distribution and evidence for isolation of demes. *Science*, 145:177–178, 1964.

[2] K. G. Ardlie. *The Frequency, Distribution and Maintenance of t-haplotypes in Natural Populations of Mice.* PhD thesis, Princeton University, 1995.

[3] K. G. Ardlie and L. M. Silver. Low frequency of mouse t haplotypes in wild populations is not explained by modifiers of meiotic drive. *Genetics*, 144:1787–1797, 1996.

[4] K. G. Ardlie and L. M. Silver. Low frequency of t-haplotypes in natural populations of house mice (*mus musculus domesticus*). *Evolution*, 52:1185–1196, 1998.

[5] A. Atlan, N. Mercot, C. Landre, and C. Montchamp-Moreau. The *sex-ratio* trait in *drosophila simulans*: Geographic distribution of distortion and resistance. *Evolution*, 51:1886–1895, 1997.

[6] D. Bruck. Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Genetics*, 43:152–58, 1957.

[7] B. Charlesworth. The evolution of lethals in the t-haplotype system of the mouse. *Proc R Soc Lond B Biol Sci*, 258:101–107, 1994.

[8] B. Charlesworth and D. L. Hartl. Population dynamics of the segregation distorter polymorphism of *drosophila melanogaster*. *Genetics*, 89:171–192, 1978.

[9] J. W . Curtsinger and M. W. Feldman. Experimental and theoretical analysis of the "sex-ratio" polymorphism of *drosophila melanogaster*. *Genetics*, 94:445–466, 1980.

[10] J. F. Dallas, B. Dod, P. Boursot, E. M. Prager, and F. Bonhomme. Population subdivision and gene flow in Danish house mice. *Molecular Ecology*, 4:311–320, 1995.

[11] L.C. Dunn. Evidence of evolutionary forces leading to the spread of lethal genes in wild populations of house mice. *Proc. Nat. Acad. Sci. USA*, 43:158–163, 1957.

[12] L.C. Dunn and Howard Levene. Population dynamics of a variant t-allele in a confined population of wild house mice. *Evolution*, 15(4):385–93, 1961.

[13] M.D. Durand. An analytical model of genetic drift in populations of lethal t-haplotypes. in preparation., 1998.

[14] M.D. Durand, K.G. Ardlie, L. Buttel, S.A. Levin, and L.M.Silver. Impact of migration and fitness on the stability of lethal t-haplotype polymorphism in house mice: A computer study. *Genetics*, 145:1093 – 1108, April 1997.

[15] A. W. F. Edwards. The population genetics of "sex-ratio" in *drosophila pseudoobscura*. *Heredity*, 16:291–304, 1961.

[16] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag, 1979.

[17] M. W. Feldman and S. P. Otto. More on recombination and selection in the modifier theory of sex-ratio distortion. *Theoretical Population Biology*, 35:207–225, 1989.

[18] M. W. Feldman and S. P. Otto. A comparative approach to the population-genetics theory of segregation distortion. *American Naturalist*, 137:443–456, 1991.

[19] M. F. Hammer and L. M. Silver. Phylogenetic analysis of the alpha-globin pseudogene-4 (Hba-ps4) locus in the house mouse species complex reveals a stepwise evolution of t haplotypes. *Mol. Biol. Evol.*, 10:971–1001, 1993.

[20] D. L. Hartl. Modifier theory and meiotic drive. *Theoretical Population Biology*, 7:168–174, 1975.

[21] D.L. Hartl and Y. Hiraizuma. Segregation distortion after fifteen years. In M. Ashburner and E. Novitski, editors, *The Genetics and Biology of* Drosophila Melanogaster. Academic Press, New York, 1976.

[22] Y. Hiraizumi, J. M. Albracht, and B. C. Albracht. X-linked elements associated with negative segretation distortion in the *sd* system of *drosophila melanogaster*. *Genetics*, 138:145–152, 1994.

[23] Y. Hiraizumi, L. Sandler, and J. F. Crow. Meiotic drive in natural populations of *drosophila melanogaster* III. populational implications of the segregation-distorter locus. *Evolution*, 14:433–444, 1960.

[24] Y. Hiraizumi and A. M. Thomas. Supressor system of segregation distortion (*sd*) chromosomes in natural populations of *drosophila melanogaster*. *Genetics*, 95:693–706, 1984.

[25] J. Jaenike. Sex-ratio drive in the *drosophila quinaria* group. *American Naturalist*, 148:237–254, 1996.

[26] S. Lenington, P. Franks, and J. Williams. Distribution of t haplotypes in natural populations of wild house mouse. *J. Mamm.*, 69:489–99, 1988.

[27] S. Lessard. The role of recombination and selection in the modifier theory of sex-ratio distortion. *Theoretical Population Biology*, 31:339–358, 1987.

[28] B.R.C. Levin, M.L. Petras, and D.I. Rasmussen. The effect of migration on the maintenance of a lethal polymorphism in the house mouse. *The American Naturalist*, 103(934):647–661, 1969.

[29] R.C. Lewontin. Interdeme selection controlling a polymorphism in the house mouse. *The American Naturalist*, 96:65–78, 1962.

[30] R.C. Lewontin and L.C. Dunn. The evolutionary dynamics of a polymorphism in the house mouse. *Genetics*, 45:705–772, 1960.

[31] U. Liberman. Theory of meiotic drive: is Mendelian segregation stable? *Theoretical Population Biology*, 10:127–132, 1976.

[32] W. Z. Lidicker. Social behavior and density regulation in house mice living in large enclosures. *Journal of Animal Ecology*, 45:677–697, 1976.

[33] W. Q. Loegering and E. R. Sears. Distorted inheritance of stem-rust resistance of Timstein wheat caused by a pollen killing gene. *Can J Genet Cytol*, 5:65–72, 1963.

[34] M. F. Lyon. Male sterility of the mouse t-complex is due to homozygosity of the distorter gene. *Cell*, 44:357–363, 1986.

[35] T. W. Lyttle. Segregation distorters. *Annu. Rev. Genet.*, 25:511–557, 1991.

[36] T. W. Lyttle. Cheaters sometimes prosper: distortion of Mendelian segregation by meiotic drive. *Trends in Genetics*, 9:205–210, 1993.

[37] G. Maffi and S. D. Jayakar. A two-locus model for polymorphism for sex-linked meiotic drive modifiers with possible applications to *aedes aegypti*. *Theoretical Population Biology*, 19:19–36, 1981.

[38] T. Morita, H. Kubota, K. Murata, M. Nozaki, C. Delabre, K. Willison, Y. Satta, M. Sakaizumi, N. Takahata, G. Gachelin, and A. Matsushiro. Evolution of the mouse t-haplotype: Recent and worldwide introgression to *mus musculus*. *Proc.Natl.Acad.Sci. USA*, 89:6851–6855, 1992.

[39] M. J. Nauta and R. F. Hoekstra. Evolutionary dynamics of spore killers. *Genetics*, 135:923–930, 1993.

[40] L. Nunney and E.M. Baker. The role of deme size, reproductive patterns and dispersal in the dynamics of t-lethal haplotypes. *Evolution*, 47(5):1342–59, 1993.

[41] E. Padieu and J. Bernet. Mode d'action des genes responsables de l'avortement de certains produits de la meiose chez l'ascomycete Podospora anserina. *Comp Rend Acad Sci Paris Ser D*, 264:2300–2303, 1967.

[42] P. A. Powers and B. Ganetsky. On the components of Segregation Distortion in *Drosophila melanogaster*. V. Molecular analysis of the *SD* locus. *Genetics*, 129:133–144, 1991.

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

[44] C. M. Rick. The tomato Ge locus: Linkage relations and geographic distribution of alleles. *Genetics*, 67:75–85, 1971.

[45] A. Robertson. Selection for heterozygotes in small populations. *Genetics*, 47:1291–1300, 1962.

[46] A. Ruvinsky, A. Polyakov, A. Agulnik, H. Tichy, F. Figueroa, and J. Klein. Low diversity of t haplotypes in the eastern form of the house mouse, *mus musculus*. *Genetics*, 127:161–168, 1991.

[47] Sandler and Golic. Segregation distortion in *drosophila*. *Trends in Genetics*, 5:181–184, 1985.

[48] R. K. Selander. Behavior and genetic variation in natural populations. *Am. Zoologist*, 10:53–60, 1970.

[49] L. M. Silver. Mouse t haplotypes. *Annual Review in Genetics*, 19:179–208, 1985.

[50] L. M. Silver. The peculiar journey of a selfish chromosome: Mouse t haplotypes and meiotic drive. *Trends in Genetics*, 9:250–254, 1993.

[51] G. J. Thomson and M. W. Feldman. Population genetics of modifiers of meiotic drive. IV. on the evolution of sex-ratio distortion. *Theoretical Population Biology*, 8:202–211, 1975.

[52] G. J. Thomson and M. W. Feldman. 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, 1976.

[53] B.C. Turner and D.D. Perkins. Spore killer, a chromosomal factor in Neurospora that kills meiotic products not containing it. *Genetics*, 93:587–606, 1979.

[54] M. van Boven. *Evolutionary Dynamics of a Selfish Gene*. PhD thesis, University of Groningen, 1997.

[55] M. van Boven, F. J. Weissing, D. Heg, and J. Huisman. Competition between segregation distorters: coexistence of "superior" and "inferior" haplotypes at the t complex. *Evolution*, 50(6):2488–2498, 1996.

[56] J. H. Werren, U. Nur, and C.-I. Wu. Selfish genetic elements. *Trends in Ecology and Evolution*, 3:297–302, 1988.

[57] C.-I. Wu and M. F. Hammer. Molecular evolution of ultraselfish genes of meiotic drive systems. In R. K. Selander, A. G. Clark, and T. S. Whitman, editors, *Evolution at the Molecular Level*, pages 177–203. Sinauer Associates, 1991.