Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 153–157, January 1985
Genetics

Interaction of selection and biased gene conversion in a multigene family

(repeated genes/molecular drive/speciation/diffusion theory/rates of evolution)

JAMES BRUCE WALSH

Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637

Communicated by James F. Crow, August 20, 1984

ABSTRACT **A model of the evolutionary dynamics of a multigene family in a finite population under the joint effects of selection and (possibly biased) gene conversion is analyzed. It is assumed that the loss or fixation of a polymorphism at any particular locus in the gene family occurs on a much faster time scale than the introduction of new alleles to a monomorphic locus by gene conversion. A general formula for the fixation of a new allele throughout a multigene family for a wide class of selection functions with biased gene conversion is given for this assumption. Analysis for the case of additive selection shows that (*i*) unless selection is extremely weak or bias is exceptionally strong, selection usually dominates the fixation dynamics, (*ii*) if selection is very weak, then even a slight conversion bias can greatly alter the fixation probabilities, and (*iii*) if both selection and conversion bias are sufficiently small, the substitution rate of new alleles throughout a multigene family is approximately the single locus mutation rate, the same result as for neutral alleles at a single-copy gene. Finally, I analyze a fairly general class of underdominant speciation models involving multigene families, concluding for these models under weak conversion that although the probability of fixation may be relatively high, the expected time to fixation is extremely long, so that speciation by "molecular drive" is unlikely. Furthermore, speciation occurs faster by fixing underdominant alleles of the same effect at single-copy genes than by fixing the same number of loci in a single multigene family under the joint effects of selection, conversion, and drift.**

There has been much recent interest in modeling the evolution of multigene families (see refs. 1–5 and references therein). The role of intergenic gene conversion in structuring multigene families has been a central issue to much of this work. With a few exceptions (6–8), most of these models have assumed no selection and no bias in gene conversion. Nagylaki and Petes (6) showed that even small amounts of conversion bias can be extremely important in the probability of fixation of a new mutant on a single chromosomal lineage. Single-locus models show that bias in conversion can effectively oppose selection in some cases (9–12), and it is of interest to inquire about their joint interactions in a multigene family. Finally, there has been much interest in the possible role of the various DNA turnover mechanisms (conversion, transposition, unequal crossing over) in promoting speciation by fixing incompatible alleles, with these turnover forces presumably overpowering the effects of selection. To address all of these concerns, I develop here a multigene family model for the joint interactions of selection, gene conversion, and genetic drift. To make an analysis possible, I assume "weak conversion," in a fashion analogous to the low-migration models Slatkin (13) and Lande (14) used to investigate geographically structured populations. In the weak-conversion limit, all of the loci in a multigene family are monomorphic (although perhaps for different alleles) except for relatively short periods of time after the introduction of a polymorphism at a single locus by gene conversion. Put another way, in this limit, the time between the actions of conversion (introducing alleles to a monomorphic locus) is very long compared with the time required for the joint effects of selection and genetic drift to fix alleles at those loci.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Formulation

Generations are discrete and nonoverlapping. The randomly mating diploid monoecious population is finite, with size characterized by its variance effective population size N_e and actual size N. Consider a gene family that is composed of a fixed number of loci n. Only two alleles are considered, a and A, and the population is initially fixed for allele A at all n loci. No mutation is assumed, so loci remain monomorphic unless intergenic gene conversion introduces the alternate allele. We assume that introduction of new alleles to monomorphic loci occurs on a much longer time scale than the subsequent loss or fixation of the introduced alleles. Under weak conversion, the fixation of a new mutant throughout a gene family occurs by first the mutant becoming fixed at a single locus and subsequently spreading from that locus to all remaining loci in the gene family.

The probability of fixation at a single locus is a standard result (15, 16), so we focus first on the probability of spread throughout a gene family given that we are initially fixed at a single locus. In the weak-conversion limit, all loci are monomorphic (although for potentially different alleles) except for very short periods of time when a polymorphism is segregating at a single locus. Genetic drift interacting with selection causes the segregating locus to become fixed, either for the introduced allele or for the original allele. After some length of time, conversion creates another polymorphic locus, and the process continues until all loci are fixed for the same allele. We model the locus-by-locus spreading process by a simple discrete time, discrete space Markov chain. The state space is $\{0, 1, \ldots, n\}$, where state i means the population consists of i loci monomorphic for allele a and $n - i$ loci monomorphic for allele A. Since we do not allow mutation, the states 0 and n are absorbing. A transition between states involves first a conversion event (the introduction into the population of a polymorphism at a monomorphic locus) and then the subsequent fixation of the introduced allele by the joint action of selection and genetic drift. In the weak-conversion limit, we can only move from state i to either state $i - 1$ or $i + 1$. We denote the transition probability from state i to state $i + 1$ by λ_i and likewise denote by μ_i the transition probability from state i to state $i - 1$.

The λ_i and μ_i completely characterize our Markov chain. Since the associated transition probability matrix is a continuant, explicit formulae for fixation probabilities and times are available (17). Specifically, the probability of fixation

starting from a single copy, $\pi(1)$, is

$$\pi(1) = 1 \Big/ \sum_{k=0}^{n-1} \rho_k \qquad [1a]$$

$$\rho_0 = 1, \qquad \rho_k = \frac{\mu_1\mu_2 \ldots \mu_k}{\lambda_1\lambda_2 \ldots \lambda_k}. \qquad [1b]$$

We now proceed to specify the general form of the λ_i and μ_i. Recall that two independent events are required for a successful state transition, the introduction by conversion of a polymorphism, and the subsequent fixation of the introduced allele. Starting from state i (the population entirely monomorphic, with allele a fixed at i loci), let λ_i^{con} be the per-generation probability that conversion introduces an a allele at a locus monomorphic for allele A, and let $u_i(1/2N)$ be the probability that allele a is fixed at this site, given it starts from a single introduced copy. Likewise μ_i^{con} and $v_i(1/2N)$ are the corresponding probabilities associated with introduction of an A allele at a locus monomorphic for allele a and the subsequent fixation of allele A at this locus. Thus, for $1 \le i \le n-1$,

$$\lambda_i = \lambda_i^{con}\, u_i(1/2N) \qquad [2a]$$

$$\mu_i = \mu_i^{con}\, v_i(1/2N). \qquad [2b]$$

Finally, denoting the initial probability of fixation of allele a at a single locus by $u_0(1/2N)$, we have our general expression for the probability that, starting from a single copy, a new mutant is fixed throughout a gene family. We denote this probability by $U(1/2N)$,

$$U(1/2N) = u_0(1/2N)\, \pi(1). \qquad [3]$$

Thus, to compute [3] we need only specify the λ_i^{con}, μ_i^{con}, $u_i(1/2N)$, and $v_i(1/2N)$.

A slight modification of the Nagylaki–Petes model allows us to obtain λ_i^{con} and μ_i^{con}. We assume that intergenic conversion occurs via a Szostak *et al.* double-strand gap repair model (18) or by asymmetric heteroduplex formation in a Meselson–Radding model (19). This restriction ensures that at most only one locus is converted per conversion event. Otherwise, the possibility exists that a single conversion event could introduce a polymorphism at two separate loci. Let γ be the probability per individual per generation that a conversion event occurs, so that $N\gamma$ conversion events occur in the population per generation. Our weak conversion assumption implies $N\gamma << 1$, so that at most a single conversion event occurs within the population.

Since γ measures all conversion events, including those between loci fixed for the same allele, only a subset of the total conversion events involve different alleles. Furthermore, even among those conversion events involving different alleles, only a subset of these result in conversion introducing a polymorphism, as the original allele may be maintained when the heteroduplex is resolved. Among those conversion events that involve the two different alleles, let p be the fraction of these events that result in allele A being converted to allele a, and likewise let q be the fraction of a converted to A (6). Note that only $(p + q)$ of the conversion events between different alleles result in an actual conversion occurring, and that, generally, $p + q < 1$. From Eqs. 3 and 4 of Nagylaki and Petes (6), for an individual in state i in which a conversion occurs, a polymorphism is introduced at an AA monomorphic locus with probability $p2i(n-i)/n(n-1)$. Likewise, a polymorphism is introduced at an aa monomorphic locus with probability $q2i(n-i)/n(n-1)$.

If $p \neq q$, conversion is said to be biased. Following Nagylaki and Petes (6), we define $r = q/p$. Equivalently, we can define α as the conditional probability that allele A is converted to allele a, given that a conversion event that introduces a polymorphism occurs. Thus, $\alpha = p/(p + q)$, with r and α being related by $r = (1 - \alpha)/\alpha$. No conversion bias occurs if $\alpha = 1/2$ $(r = 1)$, allele a is at a conversion advantage if $\alpha > 1/2$ $(r < 1)$, while allele A is at a conversion advantage if $\alpha < 1/2$ $(r > 1)$.

Putting the above results and definitions together, and noting that $p = (p + q)\alpha$, $q = (p + q)(1 - \alpha)$, we have

$$\lambda_i^{con} = N\gamma(p + q)\alpha\left[\frac{2i(n-i)}{n(n-1)}\right] \qquad [4a]$$

$$\mu_i^{con} = N\gamma(p + q)(1 - \alpha)\left[\frac{2i(n-i)}{n(n-1)}\right]. \qquad [4b]$$

For given α and γ values, the underlying molecular basis of conversion affects the rate at which actual conversions occur. We account for this by considering the effective conversion rate, which we define as $\gamma(p + q)$. A simple extension of Nagylaki and Petes (6) Eqs. 4b and 4c to include double-strand gap formation allows p and q to be computed for any arbitrary combination of interchromatid asymmetric, sister-chromatid asymmetric, and double-strand gap conversion events. Since our formulations are completely described by α and $\gamma(p + q)$, we will not consider further the formulation of p and q in terms of the underlying molecular events.

Calculation of $u_i(1/2N)$ and $v_i(1/2N)$ proceeds directly from Kimura's standard one-diallelic-locus result (15, 16) upon specification of the fitnesses. Denote the fitness of an individual with i a alleles on one haploid set and j a alleles on its other haploid set by $w(i, j)$. The genotypes AA:Aa:aa at the segregating locus have fitnesses $w(i, i)$:$w(i, i + 1)$:$w(i + 1, i + 1)$ if we start in state i and introduce allele a into a locus monomorphic for A. Since for fixed i these fitnesses are constant, the probability of fixation of allele a, $u_i(1/2N)$, is obtained by inserting these fitnesses into Kimura's general one-locus result. We compute $v_i(1/2N)$ in a similar fashion.

We expect that for many gene families the fitness of an individual depends solely on the total number of copies of allele a it carries. This introduces a great simplification in obtaining $u_i(1/2N)$ and $v_i(1/2N)$. Denote by $w(x)$ the fitness of an individual who carries allele a at a fraction x of the total loci in the gene family. In computing $u_i(1/2N)$, the genotypes AA:Aa:aa have fitnesses $w(i/n)$:$w([i + 1/2]/n)$:$w([i + 1]/n)$. If the number of loci is fairly large, we expect small changes in the total composition of allele a to produce small changes in fitness. We use this assumption for further simplification by expanding the last two fitnesses in a Taylor series about $w(i/n)$, and then normalize by dividing through by $w(i/n)$. This yields fitnesses $1{:}1 + s_i{:}1 + 2s_i$ for the genotypes AA:Aa:aa where

$$s_i = w'(i/n)/\{2nw(i/n)\}, \qquad [5]$$

where the prime denotes differentiation. The probability of fixation under additive selection is a standard result (15, 16), and from [5] we have for $1 \le i \le n-1$:

$$u_i(1/2N) \approx (1/2N)\,[4N_e s_i/\{1 - \exp(-4N_e s_i)\}] \qquad [6a]$$

$$v_i(1/2N) \approx (1/2N)\,[4N_e s_i/\{\exp(4N_e s_i) - 1\}]. \qquad [6b]$$

Finally, without loss of generality, we can take $w(0) = 1$, and we obtain $u_0(1/2N)$ from [6a] with $s_0 = w'(0)/2n$.

By inserting [4] and [6] into [1], and using this result in [3], we obtain

$$U(1/2N) = \frac{cw'(0)}{2N[1 - \exp\{-cw'(0)\}]} \Big/ \left[1 + \sum_{k=1}^{n-1} r^k \exp\left\{-c\sum_{i=1}^{k} s_i^*\right\}\right] \quad [7]$$

$$c = 2N_e/n \qquad s_i^* = w'(i/n)/w(i/n).$$

Note that [7] is independent of both the effective conversion rate $\gamma(p + q)$ and the underlying recombination map.

Results for Additive Selection

In this section, we further refine [7] under additive fitnesses, with neutral alleles as a special case obtained by letting selection vanish. Define $w(x) = 1 + sx$ as the fitness of an individual who carries allele a at a fraction x of the total sites in the gene family. If we further assume that $|s| << 1$, from [5] we have that $s_i \approx s/2n$, and [7] reduces considerably to

$$U(1/2N) = \frac{1}{2N} \frac{cs}{[1 - \exp(-cs)]} \frac{1 - \kappa}{1 - \kappa^n} \quad [8]$$

$$\kappa = r \exp(-cs), \qquad c = 2N_e/n.$$

We are most interested in the cases where conversion opposes selection—i.e., $s > 0$ and $\alpha < 1/2$ (or equivalently, $r > 1$) or $s < 0$ and $\alpha > 1/2$ ($r < 1$). Selection enters into the fixation probability in an exponential form, so that if $|cs| >> 1$, selection will dominate the fixation process unless bias is almost complete in the opposite direction. For $|cs|$ large, [8] simplifies even further. If $|cs| >> 1$, $s > 0$, and $r \exp(-cs) << 1$, then

$$U(1/2N) \approx sN_e/Nn, \quad [9a]$$

while if $|cs| >> 1$, $s < 0$, and $r \exp(-cs) >> 1$, then

$$U(1/2N) \approx (-sN_e/nN)\, r^{n-1} \exp(2N_e s). \quad [9b]$$

If we have both weak selection ($|cs| << 1$) and weak conversion bias ($\alpha = 1/2 + \beta$, $|\beta| << 1$), then $u_0(1/2N) \approx 1/2N$ and $\kappa \approx 1 - (4\beta + cs)$. If $\kappa = 1 + \xi$, where $|\xi| << 1$ and $n\xi^2 << 1$, we can approximate $(1 - \kappa)/(1 - \kappa^n)$ by $\xi/[\exp(\xi n) - 1]$. Thus, provided that $n(4\beta + cs)^2 << 1$, we have

$$U(1/2N) \approx (1/2N)(4\beta + cs), \quad \text{when } \exp[n(4\beta + cs)] >> 1 \quad [10a]$$

$$U(1/2N) \approx -(1/2N)(4\beta + cs) \exp[n(4\beta + cs)], \quad \text{when } \exp[n(4\beta + cs)] << 1 \quad [10b]$$

$$U(1/2N) \approx 1/(2Nn), \quad \text{when } n|4\beta + cs| << 1. \quad [10c]$$

Eq. **10** shows that for weak selection, small amounts of conversion bias can have a very major effect, but as seen from [**10c**] if both selection and conversion bias are sufficiently small ($|4n\beta + 2N_e s| << 1$) the probability of fixation is the same as a neutral allele with no conversion bias.

Of some interest is R, the rate of substitution of new alleles in a gene family, which we define as the expected number of alleles arising each generation that are destined to become fixed throughout the gene family. Let ν be the per locus mutation rate, so that on average in each generation $2Nn\nu$ alleles arise by mutation, each with probability $U(1/2N)$ of becoming fixed throughout the gene family, giving $R = 2Nn\nu U(1/2N)$. For [**10c**], this reduces to $R \approx \nu$, which is also the substitution rate of neutral alleles at a single locus (20). Thus, provided that selection and conversion bias are sufficiently small ($|4n\beta + 2N_e s| << 1$), the substitution rate of new alleles throughout a gene family is the same as for a single locus. Likewise, [**10a**] and [**10b**] show that biased conversion and selection can greatly increase or decrease R over that for effectively neutral alleles.

Implications for Speciation

There has been speculation on the role of gene families in speciation events, especially when differences in gene family composition are assumed to cause reduced fitness. Dover *et al.* (21, 22) have proposed that the same molecular forces, such as gene conversion, that homogenize multigene families may be able to overpower selection and allow for the spread of underdominant alleles in a gene family. This would result in the accumulation of reproductive isolation if the spread occurred in an isolated population. Dover refers to this type of speciation as molecular drive. We can gain some insight into the feasibility of such speciation models by examining the spread of underdominant alleles through a gene family. Suppose $w(i, i) = 1$, $w(i, i + 1) = w(i, i - 1) = 1 - h$, where, as before, $w(i, j)$ is the fitness of an individual with i copies of allele a on one haploid set and j copies on its other set. From this symmetry, $u_0(1/2N) = u_i(1/2N) = v_i(1/2N)$. Denote this common probability by $u(1/2N)$, which is the probability of fixation of an underdominant allele (14, 23, 24).

We compute $\pi(1)$ directly from [1] using [2] and [4] and observing that when $u_i(1/2N) = v_i(1/2N)$, these cancel in [**1b**], leaving only the effects of conversion bias, yielding

$$\pi(1) = (1 - r)/(1 - r^n). \quad [11]$$

Thus, the probability of fixation throughout a gene family of an underdominant gene fixed at a single locus is independent of the strength of selection in the weak-conversion limit. The effect of selection is as a stabilizing force, increasing the time between state transitions, which increases the expected time to fixation throughout the gene family.

It is important to see just what effect selection has on retarding the rate of spread and, hence, the time to fixation. We approximate the expected time to fixation throughout the gene family by using a diffusion approximation for our Markov chain. Let $Y(t)$ indicate the state of the Markov chain t generations after the initial fixation at the first locus (the time at which we start the chain). We have that

$$E[Y(t + 1) - Y(t) \,|Y(t) = i] = N\gamma(p + q)u(1/2N)(2\alpha - 1)2i(n - i)/n(n - 1) \quad [12a]$$

$$E[(Y(t + 1) - Y(t))^2 \,|Y(t) = i] = N\gamma(p + q)u(1/2N)2i(n - i)/n(n - 1). \quad [12b]$$

We rescale space by introducing $X_n(\tau) = Y(\tau)/n$, with time rescaled as $\tau = \delta^{-1}t$ generations, $\delta^{-1} = n^2/[2N\gamma(p + q)u(1/2N)]$. Taking the limit as δ goes to zero by letting $n \to \infty$ and $\alpha \to 1/2$, such that $n(2\alpha - 1)$ is fixed, gives the infinitesimal generator, L, for the limiting diffusion associated with our Markov chain model

$$L = (1/2)x(1 - x)d^2/dx^2 + \phi x(1 - x)d/dx, \qquad \phi = n(2\alpha - 1). \quad [13]$$

This is the same generator as for genetic drift and additive

selection and is well characterized (11, 15, 25). The expected time to fixation starting from a single fixed locus, $T(1/n)$, when $n|2\alpha - 1| << 1$ is found using Eq. 31 of Nagylaki (11) to be

$$T(1/n) \approx [2n^2/\{\gamma(p + q)\}][\sqrt{\pi}\exp(\theta) \operatorname{erf}(\sqrt{\theta})/(2\sqrt{\theta})], \quad [14a]$$

while if $n|2\alpha - 1| >> 1$ and $|2\alpha - 1| << 1$, we use Eq. 32 in ref. 11 to obtain

$$T(1/n) \approx [2n \ln(2n|2\alpha - 1|)/\{\gamma(p + q)|2\alpha - 1|\}][\sqrt{\pi}\exp(\theta) \operatorname{erf}(\sqrt{\theta})/2\sqrt{\theta}], \quad [14b]$$

where $\theta = N_e h$ and $\operatorname{erf}(x)$ is the error function (26). Eq. **14** can be further simplified when $\theta << 1$ or $\theta >> 1$. If $\theta << 1$, $\sqrt{\pi}\exp(\theta)\operatorname{erf}(\sqrt{\theta})/(2\sqrt{\theta}) \approx 1$, while it is approximately $\exp(\theta)$ when $\theta >> 1$. Thus, if $N_e h >> 1$, selection greatly increases the time to fixation over that of a neutral allele, while if $N_e h << 1$, the fixation times are essentially equivalent with those for a neutral allele.

We conclude by comparing the effectiveness of speciation induced by changes within a gene family to speciation induced by changes at many single copy loci. We assume that a small population has budded off from a larger main population, and we ask how quickly reproductive isolation accumulates. One route for the accumulation of isolation is by fixation of underdominant alleles at a number of loci, so that hybrids between the bud population and the main population are multiple heterozygotes and, hence, have reduced fitness. We contrast independent fixation at many single copy loci with the single fixation of an allele throughout a large gene family. In both situations, assume that the fitness of an individual who is heterozygous for k different loci is $(1 - h)^k$. Using Eq. 11 of Walsh (24), we can compare the time to fixation of an underdominant allele through a gene family of size n with the time required for independent fixation of underdominant alleles of the same effect at n unique loci. Under the conditions of **[14a]**, the expected time to fixation in the gene family takes $2n\nu/[\gamma(p + q)]$ times as long as independent fixation at single loci, while this ratio is $2\nu \ln(2n|2\alpha - 1|)/[\gamma(p + q)|2\alpha - 1|]$ under the conditions of **[14b]**, where ν is the per gamete mutation rate of underdominant alleles. Provided that $2\nu/[\gamma(p + q)] > 1/n$, reproductive isolation accumulates more rapidly through fixation at single copy loci than by the spread of an underdominant allele throughout a gene family.

Discussion

We have examined the probability of fixation of a new allele throughout a multigene family under the joint interaction of selection (possibly biased), gene conversion, and genetic drift. The critical assumption in our analysis is that conversion acts on a much longer time scale than the effects of selection and genetic drift. Conversion introduces a polymorphism at a monomorphic locus, and the interaction of selection and genetic drift fixes that locus before conversion introduces another polymorphism in the gene family. This weak-conversion assumption greatly simplifies our analysis. Our major conclusions are detailed below.

Selection Opposed by Biased Gene Conversion. For additive selection we showed that conversion bias is unlikely to overcome selection unless selection is quite weak. Even when the amount of per locus selection (s/n) is much less than the conversion bias (β), selection can still dominate the dynamics. From **[10]**, we see that if

$$N_e|s|/n > 2|\beta|, \quad [15]$$

then selection dominates the fixation dynamics. Thus, even if bias is much larger than the amount of per locus selection, this does not ensure that bias determines the fixation behavior. However, if alleles are effectively neutral ($|N_e s| << 1$), conversion bias greatly influences the fixation dynamics, as first demonstrated by Nagylaki and Petes (6) for chromosomal lineages.

Substitution Rates in Multigene Families Versus Single Loci. Provided that $|2N_e s + 4n\beta| << 1$, the substitution rate of alleles in a multigene family is approximately the per locus mutation rate, ν. This is the same rate as for neutral alleles at a single locus. Thus, if both selection and bias are sufficiently weak, evolution (in the sense of the appearance of new alleles that are destined to become fixed) occurs at the same rate for single copy genes as for multigene families. Nagylaki (5) shows this is also true for strictly neutral alleles with no conversion bias for arbitrary conversion rates and recombination values. Either selection or conversion bias can alter the substitution rates greatly over neutral unbiased alleles in a multigene family. Rates for single copy loci are unlikely to be greatly influenced by conversion, except under unusual circumstances (12). Bias in conversion as an important evolutionary force is mainly restricted to multigene families.

It may be possible to directly assess the importance of conversion bias in shaping gene family evolution from DNA sequence comparisons. If conversion bias is important, then the substitution rates of neutral alleles should be affected. Nagylaki and Petes (6) argued that if conversion bias is important, then most newly arising mutant alleles are likely to be at a conversion disadvantage. Thus, if conversion bias is important in gene family evolution, we expect that the third-base positions in gene families should evolve more slowly than the third-base positions of single-copy genes. Consistent with this is the observation that third-base substitution rates are reduced for class I *HLA/H-2* genes (27), a family that has been directly shown to experience conversion (28). Furthermore, we expect (from **[10b]** and **[10c]**) that this effect should be more pronounced in gene families with larger family size, because the larger family size allows for a finer discrimination of smaller conversion biases.

Fixation of Underdominant Alleles and Speciation by Molecular Drive. We have examined the fitness function $w(i, i) = 1, w(i, i + 1) = w(i, i - 1) = 1 - h$ in some detail, where $w(i, j)$ is the fitness of an individual with i copies of allele a on one haploid set and j copies on the other. In the weak-conversion limit we do not need to specify $w(i, j)$ for $|i - j| > 1$, so our analysis extends to a wide class of fitness functions of interest to speciation. Dover *et al.* (21, 22) have proposed that conversion (among other forces) allows for the fixation throughout a gene family of an underdominant allele, which results in an increase in reproductive isolation when the fixed population is crossed back to a population fixed for another allele. Our general underdominant model has the features proposed by Dover for molecular drive to be a viable speciation mechanism.

We showed that the probability of fixation of an allele throughout a gene family, given that the allele is fixed at one locus, is the same as that for a neutral allele with the same amount of conversion bias. This seems to strongly support the idea of speciation by molecular drive, provided that it is also biologically feasible (but see ref. 29). However, we also showed that while selection does not affect the probability of ultimate fixation throughout the gene family (once a single locus is fixed), it strongly affects the expected time to fixation. Since heterozygotes are at a disadvantage, selection can greatly reduce the per-generation probability of a successful change in genomic state (an increase or decrease in the number of loci fixed for allele a), which increases the expected time between successful events. The net result being that if selection is sufficiently strong ($N_e h >> 1$), it

greatly increases the expected fixation time, making this an ineffective speciation mechanism.

Another problem exists with multigene families being involved in speciation events. To see this, suppose an individual heterozygous at k loci has fitness $(1 - h)^k$, regardless of whether these heterozygous loci are in a gene family or at single copy genes. Using this fitness function, we examined which is more effective in accumulating reproductive isolation: independent fixation of underdominant alleles at single copy loci or coordinate fixation of the same allele throughout a multigene family. We showed that fixation occurs at a faster rate at n individual single copy loci than for an entire gene family. Hence, reproductive isolation accumulates more rapidly through fixation at single loci than through fixation of a single allele throughout a gene family. The reason for this is that exactly n single locus fixation events are required for the unique copy loci, but considerably more than n single locus fixations are required to fix an allele throughout a multigene family (see [14]). In the process of fixing an allele throughout a multigene family, loci that are fixed for allele a may become unfixed by conversion and refixed to allele A by drift, and these loci must be subsequently refixed for a (perhaps several times) before a is fixed in the entire family, the net result being many more than n single locus fixation events being required to fix the allele.

In conclusion, biased conversion can result in a relatively high probability of fixation throughout a gene family for an underdominant allele, but the time required to fixation can be quite long. Even if the time is sufficiently short to operate within the lifetime of a species, the accumulation of reproductive isolation by gene families is a much less efficient process than the accumulation that results at independent single copy genes.

Genomic Versus Populational Levels of Evolution. Ohta (2) describes the process of fixation of an allele throughout a gene family as a double diffusion. For any given locus, the allele must become fixed at that locus for all individuals in the population. This is diffusion at the populational level. Secondly, in addition to all loci becoming monomorphic, all loci in the gene family must become fixed for the same allele—this diffusion operates at the genomic level. Usually, evolution proceeds simultaneously at both levels, making the entire process quite complicated. In the weak conversion limit, however, these two levels of evolution are decoupled, as polymorphic loci become monomorphic on a much faster time scale than the spread of an allele to other loci.

The spread of an underdominant allele provides a useful example of how the two levels of evolution interact. If selection is strong ($N_e h >> 1$), the probability of fixation of an underdominant allele at a single locus is quite small. However, once fixation occurs at a single locus, the probability of fixation throughout the rest of the gene family depends only on the conversion bias. Fixation is thus unlikely at the populational level, but much more likely at the genomic level, especially if bias is strong. In this case, the effect of evolution at the populational level is to set the time scale for events at the genomic level. Thus, even if it is likely that the allele will eventually become fixed throughout the gene family, the time scale for a successful fixation may make such an event biologically unlikely.

We can also gain insight into the nature of genomic versus populational levels of evolution by comparing the relative drift parameters at the two levels of diffusion. If random sampling forces are strong, the deterministic force (conversion bias, selection) must be sufficiently large to overcome it. At the genomic level, the diffusion operates over the number of genes in the gene family, n, while at the populational level selection operates over the effective population size, N_e. Usually, $N_e >> n$, so that the populational level is able to discriminate smaller effects of selection than the genomic level can for the effects of bias. As an example, recall for weak additive selection and biased gene conversion that the quantity $4n\beta + 2N_e s$ determines the behavior. The amount of bias can be considerably larger than the amount of per locus selection (s/n), but provided that [15] holds, selection dominates the system. This reflects the finer discrimination for smaller deterministic forces at the populational level compared with the genomic level.

Relaxation of the Weak-Conversion Limit. It is likely that many gene families do not satisfy the weak-conversion limit conditions. What can we say about such families? The weak-conversion limit is the situation in which selection should have the least effect (compared to gene conversion) on structuring gene families. Outside of the weak-conversion limit, individuals exist in the population that are polymorphic for more than one locus. If selection is operating in such a fashion that it only depends on the number of copies of, for example, allele a these individuals carry, then variation in fitness increases as we leave the weak-conversion limit. This increased variance strengthens the effects of selection. Thus, our conclusions that even a very small amount of selection will structure a gene family hold under much more general conditions of higher conversion rates. Secondly, the weak-conversion limit provides the optimal conditions for molecular drive models of speciation, and because we have shown that even under these most favorable conditions it is an ineffective mode of speciation, it is even more unlikely under strong conversion.

Thanks to Tom Nagylaki for clarification of several important points. This work was supported by a National Institutes of Health postdoctoral fellowship.

1. Smith, G. P. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 507–513.
2. Ohta, T. (1980) *Evolution and Variation of Multigene Families* (Springer, Berlin).
3. Ohta, T. (1983) *Theor. Pop. Biol.* **23**, 216–240.
4. Ohta, T. (1983) *Genetics* **106**, 517–528.
5. Nagylaki, T. (1983) *Genetics* **106**, 529–548.
6. Nagylaki, T. & Petes, T. (1982) *Genetics* **100**, 315–337.
7. Ohta, T. (1981) *Genet. Res.* **37**, 133–149.
8. Takahata, N. (1981) *Genet. Res.* **38**, 97–102.
9. Lamb, B. C. & Helmi, S. (1982) *Genet. Res.* **39**, 199–217.
10. Hickey, D. H. (1982) *Genetics* **101**, 519–531.
11. Nagylaki, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6278–6281.
12. Walsh, J. B. (1983) *Genetics* **105**, 461–468.
13. Slatkin, M. (1981) *Evolution* **35**, 477–488.
14. Lande, R. (1979) *Evolution* **33**, 234–251.
15. Kimura, M. (1957) *Ann. Math. Stat.* **28**, 882–901.
16. Crow, J. F. & Kimura, M. (1970) *Introduction to Population Genetics Theory* (Harper & Row, New York), pp. 424–427.
17. Ewens, W. J. (1979) *Mathematical Population Genetics* (Springer, Berlin), pp. 73–74.
18. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983) *Cell* **33**, 25–35.
19. Meselson, M. S. & Radding, C. M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 358–361.
20. Kimura, M. (1969) *Nature (London)* **217**, 624–626.
21. Dover, G. A. (1982) *Nature (London)* **299**, 111–117.
22. Dover, G. A., Brown, S. D. M., Coen, E. S., Dallas, J., Strachan, T. & Trick, M. (1982) in *Genome Evolution*, eds. Dover, G. A. & Flavell, R. B. (Academic, London), pp. 347–374.
23. Kimura, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5773–5777.
24. Walsh, J. B. (1982) *Am. Nat.* **120**, 510–532.
25. Kimura, M. & Ohta, T. (1969) *Genetics* **61**, 763–771.
26. Gautschi, W. (1964) in *Handbook of Mathematical Functions*, eds. Abramowitz, M. & Stegun, I. A. (National Bureau of Standards, Washington, DC), pp. 295–309.
27. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2671–2675.
28. Mellor, A. L., Weiss, E. H., Ramachandran, K. & Flavell, R. A. (1983) *Nature* **306**, 792–795.
29. Rose, M. R. & Doolittle, W. F. (1983) *Science* **220**, 157–162.