SELECTION AND BIASED GENE CONVERSION IN A MULTIGENE FAMILY: CONSEQUENCES OF INTERALLELIC BIAS AND THRESHOLD SELECTION

JAMES BRUCE WALSH[1]

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

Manuscript received July 15, 1985
Revised copy accepted November 14, 1985

ABSTRACT

In a previous paper, I investigated the interactions in a gene family of additive selection and biased gene conversion in a finite population when conversion events are rare. Here I extend my "weak-conversion limit" model by allowing biased interallelic conversion (conversion between alleles at the same locus) of arbitrary frequency and various threshold selection schemes for rare interlocus conversion events. I suggest that it is not unreasonable for gene families to experience threshold fitness functions, and show that certain types of thresholds can greatly constrain the rate at which advantageous alleles are fixed as compared to other fitness schemes, such as additive selection. It is also shown that the double sampling process operating on a gene family in a finite population (sampling over the number of genes in the gene family and over the number of individuals in the population) can have interesting consequences. For selectively neutral alleles that experience interallelic bias, the probability of fixation at each single locus may be essentially neutral, but the cumulative effects on the entire gene family of small departures from neutrality can be significant, especially if the gene family is large. Thus, in some situations, gene families can respond to directional forces that are weak in comparison to drift at single loci.

A general theme on genome structure that is emerging from molecular biology is the ubiquity of multigene families. Genes that exist as multiple copies are subjected to additional evolutionary forces that do not operate on single loci. Because of this, it is not surprising that much attention has been paid recently to theoretical population genetics models of the evolution of multigene families (*e.g.*, OHTA 1981; NAGYLAKI 1984; WALSH 1985a; and references therein). This paper continues an examination of the joint interactions of biased gene conversion, selection, and genetic drift (WALSH 1985a). A key result from WALSH (1985a) was that even very small amounts of additive selection can overpower fairly strong interlocus conversion bias. Here, we provide an extension of our earlier analysis by (1) examining the consequences of various threshold selection schemes and (2) allowing for interallelic bias (bias in conversion between alleles at the same locus) in addition to interlocus bias.

[1] Address before September 15, 1986: Department of Genetics, University of California, Davis, California 95616.

Our main conclusions are that conversion bias can be quite important with threshold fitness functions, but that additive selection usually dominates conversion, even with fairly strong interallelic bias. For neutral alleles, however, even a very small amount of interallelic bias can be important. Finally, the specific form of the fitness function is of considerable evolutionary importance. Certain types of threshold fitness schemes are much less efficient than additive selection in fixing advantageous alleles.

To incorporate possible differences in intrallelic and interlocus bias, we start with our earlier weak-conversion limit model (WALSH 1985a), which allows only interlocus conversion events and further assumes that such events occur very infrequently. We shall keep this restriction on the rates of interlocus events, but shall place no such restriction on interallelic rates. Our motivation for building upon the weak-conversion limit model is that explicit analytic results are obtainable. Additionally, interallelic rates may be much higher than interlocus rates, although recent data from yeast (JINKS-ROBERTSON and PETES 1985) suggest that this may not always be true. In any event, the weak-conversion results provide a limiting case and suggest features that may hold in much more general models (*i.e.*, high rates of interlocus conversion), which are currently intractable.

In what follows, we first discuss the molecular evidence pertaining to interallelic *vs.* interlocus bias. Next, we formulate the generalized model for selection and then refine our analysis for two cases: additive selection (with selectively neutral alleles as a special case) and various types of threshold selection. Finally, we discuss the basic conclusions and biological implications of both our models and other models addressing various aspects of multigene family evolution.

INTERALLELIC *VS.* INTERLOCUS CONVERSION

Previous evolutionary models of gene conversion often have failed to distinguish between two fundamentally different types of conversion events: those between different alleles at the same locus (interallelic conversion) and those between alleles at different loci (interlocus conversion). Recent work suggests that key features in the enzymology of these reactions may be quite different. The general rule of thumb is that roughly 50% of all interallelic conversion events are associated with reciprocal recombination events (FINK and PETES 1984), although the two events may be partially separated for conversions occurring in the G_1 phase of mitosis (ROMAN and FARBE 1983). On the other hand, interlocus conversion events are apparently not associated with reciprocal recombination events (KLEIN and PETES 1981; KLEIN 1984; KLAR and STRATHERN 1984; JACKSON and FINK 1985). This is fortunate, for if interlocus events were frequently associated with reciprocal recombination, chromosome rearrangements would often result. This has important consequences in the homogenization by conversion of gene families whose members exist on multiple chromosomes. If conversion-induced rearrangements were frequent, newly converted alleles would often be associated with chromosomal rearrangements

and, hence, would be at a selective disadvantage, resulting in a decreased rate of homogenization.

Given differences in interlocus *vs.* interallelic conversion events, it would not be surprising if they had different biases. As noted earlier (WALSH 1983), the two currently competing classes of conversion models make opposite predictions about the direction of bias for an allele for which DNA sequence has a propensity to be nicked. In single-strand break models (MESELSON and RADDING 1975) such alleles are at a conversion advantage, whereas in double-strand break models (SZOSTAK *et al.* 1983) they are at a conversion disadvantage. Biases may also result from preferential resolution of the conversion heteroduplex, and recent data suggest that this may be the case for some interallelic events (KRAMER, KRAMER and FRITZ 1984; WHITE, LUSNAK and FOGEL 1985). Single-strand break models are expected to generate longer heteroduplexes than double-strand break models (WHITE, LUSNAK and FOGEL 1985). Thus, alleles that have a conversion bias due to preferential heteroduplex resolution would have stronger biases in single-strand break models compared to double-strand break models. At least one class of interlocus conversion events (mating type switching) requires double-strand breaks (KOSTRIKEN *et al.* 1983; EGEL, BEACH and KLAR 1984), whereas the heteroduplex data of WHITE, LUSNAK and FOGEL favors single-strand breaks for at least some interallelic conversion. Finally, there have been recent suggestions that interlocus conversion might be RNA-mediated (BERNSTEIN, MOUNT and WEINER 1983; MORZYCKA-WROBLEWSKA *et al.* 1985; DOOLITTLE 1985). All of this suggests that it is not unreasonable to assume different biases (including differences in sign) for the different classes of conversion events.

FORMULATION OF THE GENERAL MODEL

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. The gene family has a fixed size of n loci, and we consider only two allelic states, A and a. The population is initially monomorphic for allelic state A, and we wish to compute the probability of fixation for allele a, given that we introduce a single copy. We proceed by a straightforward extension of a previous model (WALSH 1985a) by allowing interallelic gene conversion in addition to interlocus conversion. We assume that conversion between different loci occurs at a sufficiently low frequency that an allele introduced by interlocus conversion to a new locus is either lost or fixed by the joint interactions of genetic drift, selection and interallelic gene conversion before the next conversion event *between* loci occurs. We refer to this class of models as weak-conversion limit models, and WALSH (1985a) can be consulted for further discussion of this assumption.

As before, we model the locus-by-locus spreading of the new allele a by a discrete time, discrete space Markov chain, with state space $\{0,1, \cdots n\}$. State i means that the population consists of i loci monomorphic for allele a and $n - i$ loci monomorphic for allelic state A. In the weak-conversion limit, the population can only move from state i to either state $i + 1$ or state $i - 1$, and

we denote the transition probabilities into these states by λ_i and μ_i, respectively. Denote by $\pi(1)$ the probability of fixation at state n, given that we start in state 1. Since the associated probability transition matrix is a continuant, $\pi(1)$ is a standard result (EWENS 1979, pp. 73–74):

$$\pi(1) = 1/\sum_{k=0}^{n-1} \rho_k \tag{1a}$$

$$\rho_0 = 1,\ \rho_k = \frac{\mu_1\mu_2 \cdots \mu_k}{\lambda_1\lambda_2 \cdots \lambda_k}. \tag{1b}$$

To compute λ_i and μ_i, note that two independent events are required for a successful state transition. First, a polymorphism must be introduced at a previously monomorphic locus by interlocus gene conversion, and second, the introduced allele must become fixed at that locus through the joint interactions of genetic drift, selection and interallelic conversion. Given that we are in state i, let λ_i^{con} be the per-generation probability that conversion introduces an a allele at a locus previously monomorphic for allele A, and let $u_i(1/2N)$ be the probability of fixation of the introduced allele, given that we start with a single copy. Define μ_i^{con} and $v_i(1/2N)$ similarly. Thus, for $1 \leq i \leq n-1$;

$$\lambda_i = \lambda_i^{con} u_i(1/2N) \tag{2a}$$

$$\mu_i = \mu_i^{con} v_i(1/2N). \tag{2b}$$

Finally, if $u_0(1/2N)$ is the probability that the introduced single copy of the new allele a initially becomes fixed at a single locus, then the probability that the allele is eventually fixed throughout the entire gene family, which we denote by $U(1/2N)$, is

$$U(1/2N) = u_0(1/2N)\pi(1). \tag{3}$$

The effects of bias and the underlying model of conversion for interlocus conversion events enter into our analysis only through λ_i^{con} and μ_i^{con}. In obtaining the probability of fixation, only the ratio of these terms ($\mu_i^{con}/\lambda_i^{con}$) matters, and this ratio is an indication of the strength of interlocus conversion bias (NAGYLAKI and PETES 1982; WALSH 1985a). A conversion interaction between A and a has several possible outcomes (MESELSON and RADDING 1975), of which we need consider only unequal conversion events. In such events the interacting (a/A) pair is converted into either (a/a) or (A/A). Let the probability that allele A is converted to allele a, given that an unequal conversion event occurred between loci, be $(1/2 + \beta)$, and likewise, let this conditional probability that a is converted to A be $(1/2 - \beta)$. β measures the conversion bias for events between different loci. Using the NAGYLAKI-PETES model for conversion, we have (for $1 \leq i \leq n-1$) $\mu_i^{con}/\lambda_i^{con} \equiv r = (1/2 - \beta)/(1/2 + \beta)$, which is independent of both i and the rate of interlocus conversion events (which we denote by γ^*).

Calculation of $u_i(1/2N)$ and $v_i(1/2N)$ proceeds directly from standard one-diallelic locus results for the probability of fixation under selection, genetic

drift and conversion bias (WALSH 1983; NAGYLAKI 1983). To apply these results, denote the fitness of an individual with i a alleles on one haploid set (gamete) and j a alleles on the other haploid set by $w(i, j)$. Starting with the population in state i, and introducing an a allele into a locus that was monomorphic for allele A gives the genotypes AA:Aa:aa at this segregating locus the fitnesses $w(i, i)$:$w(i + 1, i)$:$w(i + 1, i + 1)$. This is our most general formulation for fitnesses. In many cases, however, we might expect the fitness of an individual to simply depend on the number of copies of allele a that it carries. In this case, let $w(x)$ be the fitness of an individual that carries allele a at a fraction x of the total loci in the gene family. Further, if the number of loci is large, we might expect small changes in the total composition of allele a to produce only small changes in fitness. Under this assumption we use a first-order Taylor expansion to see that, for our segregating locus, the genotypes AA:Aa:aa have approximate fitnesses $1{:}1 + s_i{:}1 + 2\,s_i$, with

$$s_i = w'(i/n)/[2nw(i/n)] \tag{4a}$$

and prime denoting differentiation.

It remains only to incorporate the effects of bias in interallelic conversion. Interallelic conversion parameters are defined as follows. Let γ be the per-generation per-locus rate of unequal interallelic conversion events. When an unequal conversion event occurs, with probability $(1/2 + \delta)$, an allele A is converted to allele a, likewise with probability $(1/2 - \delta)$, allele a is converted to allele A. It has already been shown (GUTZ and LESLIE 1976; LAMB and HELMI 1982; WALSH 1982, 1983; NAGYLAKI 1983) that if the genotypes AA:Aa:aa have fitnesses $1{:}1 + h{:}1 + t$ and biased gene conversion parameters γ and δ, then provided that h, t, and $\gamma\delta$ are all sufficiently small so that we can ignore higher products, this locus behaves like one with no conversion bias but fitnesses $1{:}1 + h + 2\gamma\delta{:}1 + t + 4\gamma\delta$. From (4a), for the fitness function $w(x)$, we compute $u_i(1/2N)$ using fitnesses $1{:}1 + \xi_i{:}1 + 2\xi_i$, where

$$\xi_i = 2\gamma\delta + w'(i/n)/[2nw(i/n)] \tag{4b}$$

where the prime denotes differentiation. Thus, we have reduced our system to an equivalent one with additive selection and can use standard results (KIMURA 1957) to obtain

$$u_i(1/2N) \approx [1/2N][4N_e\xi_i/(1 - \exp\{-4N_e\xi_i\})] \qquad 0 \le i \le n - 1 \tag{5a}$$

and likewise

$$v_i(1/2N) \approx [1/2N][4N_e\xi_i/(\exp\{4N_e\xi_i\} - 1)] \qquad 0 \le i \le n - 1. \tag{5b}$$

Using (5) and (1) and recalling that $\mu_i^{con}/\lambda_i^{con} = r$, we obtain

$$\pi(1)^{-1} = 1 + \sum_{k=1}^{n-1} r^k \exp\{-k\theta\} \exp\{-c\sum_{i=1}^{k} \tilde{s}_i\} \tag{6}$$

where

$$c = 2N_e/n, \qquad \theta = 8N_e\gamma\delta, \qquad \tilde{s}_i = w'(i/n)/w(i/n).$$

Equation (6) shows the decomposition of the effects of interlocus conversion bias (r), interallelic conversion bias (θ) and selection ($\tilde{s}_i$). From (6) we see that selection in the earlier states is given more weight than the same amount of selection occurring in later states. We shall see the implications of this when we examine threshold selection.

RESULTS

Additive selection

We can further refine (6) by assuming additive selection, with neutrality as a special case. Define $w(x) = 1 + sx$ as the fitness of an individual with allele a at a fraction x of the total sites in the gene family. Provided $|s| \ll 1$, $s_i \approx s/2n$, and (6) reduces to

$$U(1/2N) \approx \frac{1}{2N} \frac{cs + \theta}{1 - \exp\{-(cs + \theta)\}} \frac{1 - \xi}{1 - \xi^n} \tag{7}$$

where $\xi = r \exp\{-(cs + \theta)\}$, and θ and c as defined in (6).

The case of additive selection and interlocus biased conversion has been examined elsewhere (WALSH 1985a), so we focus here on three cases: (1) selectively neutral alleles with only interallelic bias ($s = \beta = 0$, $\delta \neq 0$); (2) selectively neutral alleles with both interallelic and interlocus bias ($s = 0$; δ, $\beta \neq 0$); and (3) the general case of additive selection and both interallelic and interlocus bias (s, δ, $\beta \neq 0$).

Case 1.1: Neutral alleles, no interlocus conversion bias: This case provides a convenient starting point and also provides interesting insight into one of the more subtle features of the double-diffusion operating in gene family evolution (OHTA 1981). From (7) it follows that the probability of fixation of a single new mutant through the gene family is

$$U(1/2N) \approx [1/2N][\theta/(1 - \exp\{-n\theta\})] \tag{8}$$

thus,

$$U(1/2N) \approx 1/[2Nn] \quad \text{if} \quad |n\theta| \ll 1, \tag{9a}$$

$$U(1/2N) \approx 4\gamma\delta[N_e/N] \quad \text{if} \quad n\theta \gg 1, \text{ and} \tag{9b}$$

$$U(1/2N) \approx [-\theta/2N]\exp\{n\theta\} \quad \text{if} \quad n\theta \ll -1. \tag{9c}$$

These provide an interesting contrast to one-locus results (WALSH 1983; NAGYLAKI 1983). For single-locus fixation probabilities, $\theta = 8N_e\gamma\delta$ determines the behavior, whereas for the fixation probabilities for a gene family of size n, $n\theta$ is the critical parameter. At a single locus, allele a behaves essentially as a neutral allele if $|\theta| \ll 1$. However, from (9a) we see that for fixation throughout a gene family, a behaves as a neutral allele only if $n|\theta| \ll 1$. Thus, if $|\theta| \ll 1$, but $n|\theta| \gg 1$, allele a behaves essentially neutrally for each single-locus fixation event, but the cumulative effects of the very small departures from neutrality at each locus provide for potentially large departures from strict neutral expectations [*i.e.*, (9a)] when the entire gene family is considered. This

is a reflection of the nature of the double diffusion operating on the gene family, where sampling occurs both over the gene family (sampling over n, gene family members) and the population (sampling over N_e individuals). Even in the weak-conversion limit where these two sampling processes are uncoupled (WALSH 1985a), we still see that, in some cases (such as interallelic bias), the total sampling process behaves as if it had size $N_e n$, allowing for a much finer discrimination of deterministic forces than is obtained by either sampling process separately.

Case 1.2: Neutral alleles, both interlocus and interallelic bias: In this section we examine the relative importance of interallelic *vs.* interlocus bias for selectively neutral alleles when interlocus conversion rates are low. From (7) we have that

$$U(1/2N) \approx \frac{1}{2N}\,\frac{\theta}{1-\exp\{-\theta\}}\,\frac{1-r\exp\{-\theta\}}{1-r^n\exp\{-n\theta\}}. \tag{10}$$

If $r\exp\{-\theta\} \ll 1$ and $\theta \gg 1$,

$$U(1/2N) \approx 4\gamma\delta[N_e/N] \tag{11a}$$

whereas if $r\exp\{-\theta\} \gg 1$ and $\theta \ll -1$,

$$U(1/2N) \approx [-\theta/2N]\exp\{\theta\}[1/r\exp\{-\theta\}]^{n-1}, \tag{11b}$$

which is bounded above by $[-\theta/2N]\exp\{\theta\}$, for $r\exp\{-\theta\} \gg 1$. More generally, we see directly from (10) that since the effects of interallelic bias (θ) enter as exponential terms, whereas interlocus bias enters as $r \equiv (1-2\beta)/(1+2\beta)$, very large amounts of interlocus biases ($\beta = 1/2 + \epsilon$, $0 < -\epsilon \ll 1$; or $\beta = -1/2 + \epsilon$, $0 < \epsilon \ll 1$) are required to overcome interallelic biases when $8N_e\gamma|\delta| \equiv |\theta| \gg 1$. Thus, in the weak-conversion limit, the effects of conversion bias acting between alleles at the same locus are generally more important than the effects of bias acting between loci. The assumption of weak interlocus conversion rates implies $2N\gamma^* \ll 1$, where N is the actual population size and γ^* is the rate of interlocus conversion. When $|\theta| \gg 1$, for our weak conversion assumption to still hold, we require that $\gamma \gg \gamma^*$, so our above result is intuitively obvious. Suppose, however, that $|\theta| \ll 1$ and, further, that bias between loci is also weak (*e.g.*, $|\beta| \ll 1$), then we have

$$U(1/2N) \approx [1/2N]4[\beta + 2N_e\gamma\delta] \quad \text{if} \quad \exp\{n(4\beta+\theta)\} \gg 1 \tag{12a}$$

$$U(1/2N) \approx [1/2N](-4)[\beta + 2N_e\gamma\delta]\exp\{n(4\beta+\theta)\} \quad \text{if} \quad \exp\{n(4\beta+\theta)\} \ll 1, \tag{12b}$$

$$U(1/2N) \approx [1/2Nn] \quad \text{if} \quad n|4\beta+\theta| \ll 1. \tag{12c}$$

From (12c) we see that bias has no effect if $4n|\beta + 2N_e\gamma\delta| \ll 1$. When $4n|\beta + 2N_e\gamma\delta| \gg 1$, interlocus bias dominates when $|\beta| \gg 2N_e\gamma|\delta|$, whereas interallelic bias dominates when the inequality is reversed. Thus, for small populations and/or very low rates of interallelic conversion (so that $N_e\gamma \ll 1$), the

effects of interlocus bias are more important than the effects of interallelic bias in determining the fixation dynamics of selectively neutral alleles.

Case 1.3: General case: We see from direct examination of (7) that, if $|cs + \theta| \gg 1$, then the joint effects of selection and interallelic bias overpowers all but the most extreme interlocus bias (when we are in the weak-conversion limit). For this case, additive selection dominates provided that $|s| > 4n\gamma|\delta|$, whereas interallelic bias dominates if the inequality is reversed. Since γ is expected to be small for most nonfungal systems ($<10^{-3}$, LAMB 1984), even very small amounts of additive selection can overpower fairly large interallelic biases (δ). Further, for many structural RNA and protein-coding families, n is moderate, usually <1000 and often <20 (*e.g.*, LONG and DAWID 1980; OLD and WOODLAND 1984; KAFATOS 1983; MOORE, CONKLING and GOODMAN 1982; FYRBERG *et al.* 1980; PIATIGORSKY 1984). With the expectation of rDNA and histones in a few species, most gene families with large n ($\gg$1000) have unknown function (reviewed by BOUCHARD 1982) and simply may be sequences generated and maintained by genomic-level processes. Selection acting on such families most likely acts directly on copy number and not on specific sequence variants, in which case the selection models of CHARLESWORTH and CHARLESWORTH (1983) and LANGLEY, BROOKFIELD and KAPLAN (1983), and the nonselective model of WALSH (1985b), are more appropriate.

Suppose that both selection and interallelic bias are weak, *i.e.*, $|cs| \ll 1$ and $|\theta| \ll 1$, and that interlocus bias is also fairly weak ($|\beta| \ll 1$). The critical parameter in this situation is $4[\beta + N_e(2\gamma\delta + s/2n)]$, which we denote by Λ in what follows. If $n\ |\Lambda| \ll 1$, then the probability of fixation for allele a is approximately that for a selectively neutral allele with no conversion bias (*i.e.*, $1/[2Nn]$). When $n|\Lambda| \gg 1$, the sign of Λ determines the behavior, with

$$U(1/2N) \approx [2N_e/N][2\gamma\delta + \beta/N_e + s/2n] \quad \text{when} \quad \exp\{n\Lambda\} \gg 1 \qquad (13a)$$

$$U(1/2N) \approx [1/2N][-\Lambda]\exp\{n\Lambda\} \quad \text{when} \quad \exp\{n\Lambda\} \ll 1. \qquad (13b)$$

From (13) we find that selection dominates both forms of conversion bias provided

$$|s| > 4n\gamma|\delta| + 2n|\beta|/N_e. \qquad (14)$$

From which we again see that a small amount of additive selection can be quite powerful in structuring a gene family, even in the face of fairly strong bias (β and δ). Equation (14) is also of interest for departures from the weak-conversion limit. Two features are introduced by such departures: segregation at multiple sites and bias from interlocus conversion events acting in addition to interallelic bias to influence the fixation probabilities at single loci. We have suggested elsewhere that the addition of segregation at multiple sites strengthens the effects of additive selection, by providing the population with an increased variance in fitness upon which additive selection can act (WALSH 1985a). From our above analysis, the effects of interlocus bias acting at single loci is unlikely to overpower additive selection, as it would enter in a very similar fashion as interallelic bias and, thus, would allow weak additive selection

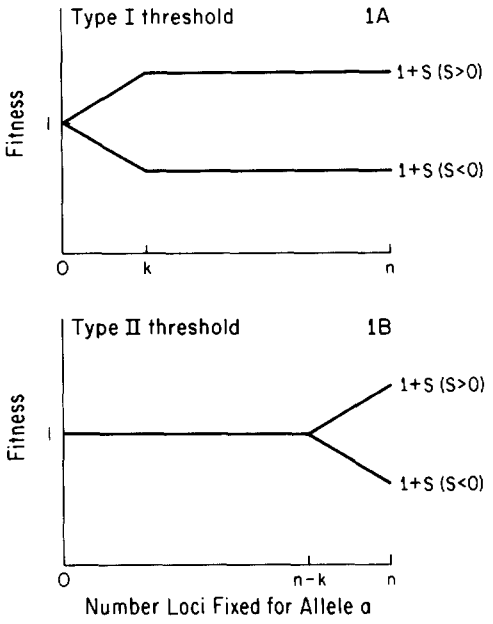

FIGURE 1.—The two types of threshold fitness functions that we consider here. The horizontal axis is the number of loci that are fixed for allele a, with n loci total in the gene family, and the vertical axis is the individual fitness as a function of number of loci fixed for a. A, a type I threshold, where the allele a has fitness effect that increases with copy number until k of the loci are fixed for a, after which the fitness is unaltered by increased a copy number. B, a type II threshold, where, initially, a has no effect on fitness, and fitness effects are apparent only after $n - k$ of the loci are fixed for a. Under both schemes, in the linear part of each fitness function the substitution of an a allele at a single locus changes fitness by s/k. See the text for possible biochemical mechanisms that might generate these types of thresholds.

to overpower even fairly strong interlocus bias. SLATKIN (1985) has examined a deterministic model of the effects of conversion bias and selection, which can be regarded as the opposite case from the weak-conversion limit. In his analysis, the relative evolutionary forces (selection and interlocus conversion) enter in such a way that selection dominates (in our notation), provided $|s| > 4n\gamma^*|\beta|$, which agrees with our above assertion, as can be seen by replacing γ (the rate of interallelic conversions) by γ^* (the rate of interlocus conversions) in our equation (14).

Threshold selection

The motivation behind examining threshold selection is two-fold. First, previous analysis has been limited to additive selection, and threshold selection provides an alternative formulation to investigate the effects of selection. Second, as we discuss below, we might expect that genes which exist in multiple copies may be kinetically buffered, resulting in threshold fitness functions. Formulation of different types of threshold schemes allows us to assess the evolutionary consequences of such buffering.

Figure 1 illustrates the two types of thresholds considered here. Figure 1A depicts what we shall refer to as a "type I" threshold. Here, the initial presence

of the new allele a has a fitness effect that increases with a copy number until a threshold level is reached, beyond which addition of further copies of a does not effect fitness. As an example of a biochemical pathway that could generate such a threshold, imagine that the gene product from allele a alters a rate-limiting step. However, once the concentration of the gene product of a becomes sufficiently high (relative to A), that step is no longer rate limiting. If the selectable difference is due to the effects of the rate-limiting step, this can generate a type I threshold.

Figure 1B illustrates the second type of threshold (type II) considered. Here, allele a behaves as a selectively neutral allele until a critical number of loci containing a is reached, after which the fitness effects of a increase with a copy number. A gene family producing more product than is required can show this type of threshold. At low concentrations of the allele a product, the higher concentrations of the A product mitigates the effects of a, be they advantageous or deleterious. Only after the copy number of a reaches a critical threshold will the concentration of a product be sufficiently high to alter fitness. Genes which show dosage compensation (so that the same amount of product is made even though different number of genes may be involved) are likely to show this type of threshold. rDNA gene families (ENDOW 1980) and, more recently, U1 snRNA gene families (MANGIN, ARES and WEINER 1985) have been shown to be dosage compensated to some degree, so that compensation may be a general feature of many gene families.

It is of some interest to compare the evolutionary consequences of different patterns of selection (additive, type I/type II thresholds) when the total amount of selection experienced by an allele going to fixation is the same for all three patterns. In the APPENDIX, we compute $\pi(1)$ for both types of threshold schemes examined here. We can obtain the general conclusions by considering two cases separately. First, we examine the probability of fixation with no interallelic or interlocus bias; then we examine the interactions of conversion bias and threshold selection.

Case 2.1: Threshold selection, no bias ($\delta = \beta = 0$): For a type I threshold from (A.1) and (A.5), we have

$$U_I(1/2N) \approx \frac{1}{2N}\,\frac{c^*s}{1 - \exp\{-c^*sk\} + c^*s(n-k)\exp\{-c^*sk\}} \tag{15}$$

where $c^* = 2N_e/k$, and k is the inflection point of the threshold fitness function (see Figure 1A). If $|c^*s| \gg 1$, (15) simplifies further to

$$U_I(1/2N) \approx [s/k][N_e/N] \quad \text{for} \quad c^*s \gg 1 \tag{16a}$$

$$U_I(1/2N) \approx [1/2N][1/(n-k)]\exp\{2N_es\} \quad \text{for} \quad c^*s \ll -1. \tag{16b}$$

For a type II threshold scheme (Figure 1B), we set the inflection point at $(n - k)$, so that the total amount of selection for fixed s is the same as for a type I threshold (with inflection point at k). From (A.4)–(A.6) we obtain

$$U_{II}(1/2N) \approx \frac{1}{2N}\,\frac{c^*s}{1 - \exp\{-c^*sk\} + c^*s(n-k)}, \tag{17}$$

which simplifies further when $|c^*s| \gg 1$ to

$$U_{II}(1/2N) \approx [1/2N][1/(n-k)] \quad \text{for} \quad c^*s \gg 1 \tag{18a}$$

$$U_{II}(1/2N) \approx [1/2N][-c^*s]\exp\{2N_e s\} \quad \text{for} \quad c^*s \ll -1. \tag{18b}$$

For $c^*s \gg 1$, from (16a) and (18a) we have $U_I(1/2N)/U_{II}(1/2N) \approx 2[N_e s/k][n-k]$, so that for alleles at a selective advantage, the underlying threshold scheme makes a large difference, even though alleles becoming fixed experience the same total amounts of selection under either scheme. More generally, it is easy to show from (15) and (17) that if $s > 0$, fixation occurs with a higher probability under a type I threshold, whereas if $s < 0$, fixation is easier under a type II threshold.

It is also of interest to compare both type I and type II threshold schemes to additive selection. Assume that the fitness of a completely homozygous *aa* individual (*i.e.*, all n loci are *aa*) is $1 + s$ for all three fitness schemes. The total amount of selection is the same under all three schemes, but different probabilities of fixation result. Denoting the probability of fixation under additive selection by $U_A(1/2N)$, for advantageous mutants ($s > 0$) we have that

$$U_A(1/2N)/U_{II}(1/2N) \approx 2[N_e s/n][n-k] \quad \text{for} \quad c^*s \gg 1, \tag{19a}$$

$$U_I(1/2N)/U_A(1/2N) \approx n/k \quad \text{for} \quad c^*s \gg 1. \tag{19b}$$

Thus, the form of fitness function can have important consequences for the probability of fixation. A type I threshold is slightly more efficient than additive selection for fixing advantageous mutants, whereas compared to both additive selection and type I thresholds, type II thresholds have a much lower probability of fixation for advantageous alleles. This points out an important feature of gene family evolution: certain types of fitness functions may greatly reduce the ability of a gene family to respond to natural selection in the weak-conversion limit (that is, when interlocus conversion events are rare). The reason for this is that, under a type II threshold, the advantageous allele is fixed and maintained at the first $(n-k)$ loci simply by drift. Under either a type I or additive model, however, the allele is initially under selection, resulting in a much higher probability of both becoming established and persisting at the first few loci, which then allows the allele to spread through the rest of the gene family.

Case 2.2: Conversion bias and threshold selection: The above analysis points out that, even in the absence of conversion bias, a type II threshold can have a greatly reduced rate of adaptive evolution in the weak-conversion limit compared with either additive selection or a type I threshold. Not surprisingly, bias in either (or both) interallelic and interlocus conversion events can be critical in determining if a selectively advantageous allele with a type II threshold can become fixed. Conversion bias is unimportant if $(n-k)4|\beta + 2N_e\gamma\delta| \ll 1$, and in this case, the probability for fixation under a type II threshold is still given by (18). When bias is important, two cases are of interest: (1) when a has a conversion advantage and (2) when a is at a conversion disadvantage. We shall focus solely on selectively advantageous alleles (so that we assume c^*s

$\gg 1$ in what follows). If allele a has a nontrivial conversion advantage (*i.e.*, $(n - k)4|\beta + 2N_e\gamma\delta| \gg 1$), then

$$U_{II}(1/2N) \approx [1/2N]4[\beta + 2N_e\gamma\delta]. \tag{20}$$

Comparing (20) with (18a) shows that, if the newly arising allele has a sufficiently strong conversion advantage, the constraining effects of the threshold can be partially overcome. However, if a is at a conversion disadvantage, so that $(n - k)4[\beta + 2N_e\gamma\delta] \ll -1$, then we can use the first term in (A.4) to place an upper limit on the probability of fixation (this limit being independent of s):

$$U_{II}(1/2N) < -[1/2N]4[\beta + 2N_e\gamma\delta]\exp\{4(n - k)[\beta + 2N_e\gamma\delta]\}. \tag{21}$$

Comparing (21) and (18a) shows that bias greatly increases the constraining effects of the threshold. NAGYLAKI and PETES (1982) have suggested that, on average, a newly arising allele is more likely to be at a conversion disadvantage than at a conversion advantage, so that (21) may be much more important than (20). If this is the case, then conversion bias generally reduces the ability of alleles experiencing certain threshold fitness functions to respond effectively to selection, further compounding the ineffectiveness of such alleles compared to alleles experiencing other types of fitness functions.

DISCUSSION

We have extended our previous analyses (WALSH 1985a) on the interaction of selection and biased gene conversion in a multigene family by including bias in conversion between alleles at the same locus and by examining the implications of threshold fitness functions. It should be stressed that the analytic results presented hold exactly only in the weak-conversion limit case, where the rate of interlocus conversion events is very low (WALSH 1985a). However, as suggested below, we feel that the basic implications of the weak-conversion limit results are still valid in far more general settings.

One of the major conclusions from our previous weak-conversion limit analysis was that even small amounts of additive selection are likely to overpower stronger interlocus bias (*i.e.*, even when $|\delta| > |s|$). The inclusion of interallelic bias does not alter this, and the analysis at the end of *Case 1.3* suggests that this conclusion is robust in the sense that it holds outside the weak conversion limit. The deterministic equilibrium results of SLATKIN (1985) provide an important comparison to our weak-conversion results and support our conclusion that weak additive selection can have critical roles in structuring gene families.

Our analysis of the effects of threshold selection in the weak-conversion limit suggests that certain types of kinetic buffering of the products from multigene families (type II thresholds, see Figure 1B) can greatly decrease the rate of adaptive evolution in the weak-conversion limit. The reason for the reduced efficiency in fixing adaptive alleles for these thresholds is that such alleles must reach a certain threshold purely by genetic drift and, perhaps, in the face of conversion bias against them. The importance of this constraint outside the weak-conversion limit depends on both the nature of the threshold and the

variation in the total number of loci containing a that can be produced by purely nonselective forces (*i.e.*, conversion and drift). If the processes generating variation are such that individuals with a copy number above the selective threshold are routinely produced even when starting from a single new variant, then this constraint is unimportant. If, however, a considerable amount of genetic drift is required to cross the threshold, this constraint can be important even outside the weak-conversion limit.

Thus, the general form of the fitness function acting on a particular gene family can be as important as the total amount of selection. The type of gene families that experience the forms of selection considered here are most likely those families that exist because large amounts of their product are needed by the cell. What can we say about the relative likelihoods of additive *vs.* type I *vs.* type II thresholds for such families? If excess amounts of product are made, then we would expect type II thresholds. If rates are critical, but excess amounts of product are made, then we might see type I thresholds. If both amount and timing are critical, additive or nearly additive (*i.e.*, strictly monotonic) selection can result. An important caveat for all cases is that, although we have assumed that the number of loci in the gene family remains constant, this is clearly not the case for many families. Genes existing in tandem arrays (*i.e.*, histones, rRNA) are subjected to unequal crossing over, resulting in an amplification and deletion in the number of loci (TARTOF 1974; SMITH 1974). Such genes may simply just increase the number of loci containing functional copies to compensate for defective copies, and the analysis of such systems is an important area for future research.

Finally, we have found that the double diffusion operating on a gene family (sampling over gene family copy number n and population size N_e) can have interesting consequences. For selectively neutral alleles experiencing only interallelic conversion bias, the amount of bias may be sufficiently small that each single locus fixation event is essentially neutral, but the cumulative effects of very small departures from neutrality at each locus can provide potentially large departures from neutrality for fixation throughout the entire gene family. This is especially true for large gene families. The implication is that gene families can be more responsive to weaker deterministic forces than can single loci. Outside of the weak-conversion limit we expect this to be even more true, because the weak-conversion limit uncouples the two sampling processes, whereas in general they act simultaneously.

A useful way of visualizing the various evolutionary forces acting on gene families is to consider evolution acting at two different (but not necessarily independent) levels: population-level evolution and genomic-level evolution. Evolution at the genomic level is the introduction of a gene family member to a new site, or the replacement of one variant at that site by another—departures from normal Mendelian segregation caused by molecular interactions. Population-level evolution is the subsequent loss of fixation of the introduced variant at that site throughout the population. At the genomic level, the sampling process is operating over the number of genes in the gene family (which may be changing), and the sampling forces can have directional components.

Examples of genomic level forces include interlocus gene conversion, unequal crossing over, gene amplification, transposition and insertion of reverse-transcribed cDNAs. At the population-level, the sampling occurs over N_e, the variance effective population size. Selection and interallelic conversion bias provide examples of directional forces operating at the population level. Clearly, in large populations both levels of evolution are acting simultaneously. The importance of the sampling at both levels is that it leads to homogenization of gene family members without having to invoke any other forces such as selection (SMITH 1974; BLACK and GIBSON 1974; OHTA 1981).

Many previous models of gene family evolution have examined the important case of the equilibrium produced by homogenization due to drift at both levels being opposed by the introduction of new variants by mutation, assuming that no directional forces (*i.e.*, selection, conversion bias) are operating. The models treated here and elsewhere (NAGYLAKI and PETES 1982; WALSH 1985a) deal with the transient dynamics of gene families in the absence of mutation, but allow for directional forces in addition to drift. By examining probabilities of fixation of new alleles, we can obtain a feel for the importance of directional forces in structuring gene families, which complements the fuller equilibrium analysis of the neutral models. By introducing mutation into our models and by suitable diffusion approximations of the associated Markov chain [given by (2)], we could examine the equilibrium properties of two-allele weak-conversion limit models, but a multiple allele approach is preferred to allow comparisons with the strictly neutral models. Deterministic (*i.e.*, infinite population size) equilibrium models with selection, bias and mutation are becoming available (WALSH 1984 and unpublished results; SLATKIN 1985), and these should further allow us to assess the importance of various evolutionary forces in structuring gene families.

The picture emerging from the theoretical models and molecular studies is that drift (at both levels), genomic-level directional forces and selection can all play important roles in shaping gene families. First, it appears that for many gene families extra gene copies are often generated by a variety of molecular events—gene amplification (SCHIMKE 1984), unequal crossing over (TARTOF 1974) and insertion of processed cDNA copies (WALSH 1985b), to name a few. Such copies are often unlikely to be under selective pressure to maintain specific sequences, but can be homogenized to various extents by genomic-level drift and, possibly, by directional forces. The models of OHTA and others allow for assessment of the amount of homogenization of such selectively neutral sequences. Our results suggest that when selection is operating on a gene family, it quite often dominates genomic-level processes, placing constraints on the allowable divergence of active sequences, provided that we hold copy number fixed. Likewise, among selectively equivalent sequences, genomic-level forces are quite important. In this fashion we can imagine that our allelic states a and A are composed of collections of selectively equivalent alleles and that genomic-level forces play critical roles in structuring the composition of alleles within each state. Thus, although it is extremely unlikely that genomic-level forces can overpower the effects of selection to the extent that they drive

phenotypic evolution (DOVER 1982), such forces nevertheless play important, if not critical, roles in structuring the genome within the often weak constraints imposed by phenotypic evolution.

I thank SUE JINKS-ROBERTSON and TOM PETES for useful discussions on conversion; MICHAEL TURELLI, TOM NAGYLAKI, MONTY SLATKIN, BRUCE WEIR and TOMOKO OHTA for critical comments on earlier versions; and MONTY SLATKIN for sharing unpublished results. This work was supported by a National Institutes of Health postdoctoral fellowship.

LITERATURE CITED

BERNSTEIN, L. B., S. M. MOUNT and A. M. WEINER, 1983 Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell **32:** 461–472.

BLACK, J. A. and D. GIBSON, 1974 Neutral evolution and immunoglobulin diversity. Nature **250:** 327–328.

BOUCHARD, R. A., 1982 Moderately repetitive DNA in evolution. Int. Rev. Cytol. **76:** 113–193.

CHARLESWORTH, B. and D. CHARLESWORTH, 1983 The population dynamics of transposable elements. Genet. Res. **42:** 1–27.

DOOLITTLE, W. F., 1985 RNA-mediated gene conversion? Trends Genet. **1:** 64–65.

DOVER, G., 1982 Molecular drive: a cohesive mode of species evolution. Nature **299:** 111–117.

EGEL, R., D. H. BEACH AND A. J. S. KLAR, 1984 Genes required for initiation and resultion of mating-type switching in fission yeast. Proc. Natl. Acad. Sci. USA **81:** 3481–3485.

ENDOW, S. A., 1980 On ribosomal gene compensation in *Drosophila.* Cell **22:** 149–155.

EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, New York.

FINK, G. R. and T. D. PETES, 1984 Gene conversion in the absence of reciprocal recombination. Nature **310:** 728–729.

FYRBERG, E. A., K. L. KINDLE, N. DAVIDSON and A. SODJA, 1980 The actin genes of *Drosophila*: a dispersed multigene family. Cell **19:** 365–376.

GUTZ, H. and H. F. LESLIE, 1976 Gene conversion: a hitherto overlooked parameter in population genetics. Genetics **83:** 861–866.

JACKSON, J. A. and G. R. FINK, 1985 Meiotic recombination between duplicated elements in *Saccharomyces cerevisiae.* Genetics **109:** 303–332.

JINKS-ROBERTSON, S. and T. D. PETES, 1985 High-frequency meiotic gene conversion between repeated genes on nonhomologous chromosomes in yeast. Proc. Natl. Acad. Sci. USA **82:** 3350–3354.

KAFATOS, F. C., 1983 Structure, evolution, and developmental expression of the chorion multigene families in silkmoths and *Drosophila.* pp. 33–61. In: *Gene Structure and Regulation in Development.* Alan R. Liss, New York.

KIMURA, M., 1957 Some problems of stochastic processes in genetics. Ann. Math. Stat. **28:** 882–901.

KLAR, A. J. S. and J. N. STRATHERN, 1984 Resolution of the recombination intermediates generated during yeast mating type switching. Nature **310:** 744–748.

KLEIN, H. L., 1984 Lack of association between intrachromosomal gene conversion and reciprocal exchange. Nature **310:** 748–753.

KLEIN, H. L. and T. D. PETES, 1981 Intrachromosomal gene conversion in yeast. Nature **289:** 144–148.

KOSTRIKEN, R., J. N. STRATHERN, A. J. S. KLAR, J. B. HICKS and F. HERRON, 1983 A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. Cell **35:** 167–174.

KRAMER, B., W. KRAMER and H.-J. FRITZ, 1984 Different base/base mismatches are corrected with different efficiencies by methyl-directed DNA mismatch-repair system of *E. coli*. Cell **38:** 879–887.

LAMB, B. C., 1984 The properties of meiotic gene conversion important in its effect on evolution. Heredity **53:** 113–138.

LAMB, B. C. and S. HELMI, 1982 The extent to which gene conversion can change allele frequencies in populations. Genet. Res. **39:** 199–217.

LANGLEY, C. H., J. F. Y. BROOKFIELD and N. KAPLAN, 1983 Transposable elements in mendelian populations. I. A theory. Genetics **104:** 457–471.

LONG, E. O. and I. B. DAWID, 1980 Repeated genes in Eukaryotes. Annu. Rev. Biochem. **49:** 727–764.

MANGIN, M., M. ARES AND A. M. WEINER, 1985 U1 small nuclear RNA genes are subjected to dosage compensation in mouse cells. Science **229:** 272–275.

MESELSON, M. S. and C. M. RADDING, 1975 A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72:** 358–361.

MOORE, D. D., M. A. CONKLING and H. M. GOODMAN, 1982 Human growth hormone: a multigene family. Cell **29:** 285–286.

MORZYCKA-WROBLEWSKA, E., E. U. SELKER, J. N. STEVENS and R. L. METZENBERG, 1985 Concerted evolution of dispersed *Neurospora crassa* 5S RNA genes: pattern of sequence conservation between allelic and nonallelic genes. Mol. Cell. Biol. **5:** 46–51.

NAGYLAKI, T., 1983 Evolution of a finite population under gene conversion. Proc. Natl. Acad. Sci. USA **80:** 6278–6281.

NAGYLAKI, T., 1984 Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA **81:** 3796–3800.

NAGYLAKI, T. and T. D. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics **100:** 315–337.

OHTA, T., 1981 *Evolution and Variation in Multigene Families*. Springer-Verlag, New York.

OHTA, T., 1985 A model for duplicative transposition and gene conversion for repetitive DNA families. Genetics **110:** 513–524.

OLD, R. W. and H. R. WOODLAND, 1984 Histone genes: not so simple after all. Cell **38:** 624–626.

PIATIGORSKY, J., 1984 Lens crystallins and their gene families. Cell **38:** 620–621.

SCHIMKE, R. T., 1984 Gene amplification in cultured animal cells. Cell **37:** 705–713.

SLATKIN, M., 1986 Interchromosomal biased gene conversion, mutation and selection in a multigene family. Genetics **112:** 681–698.

SMITH, G. P., 1974 Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. **38:** 507–513.

TARTOF, K., 1974 Unequal mitotic sister chromatid exchange as the mechanism for ribosomal gene magnification. Proc. Natl. Acad. Sci. USA **71:** 1272–1276.

WALSH, J. B., 1982 Rate of accumulation of reproductive isolation by chromosome rearrangements. Am. Nat. **120:** 510–532.

WALSH, J. B., 1983 Role of biased gene conversion in one-locus neutral theory and genome evolution. Genetics **105:** 461–468.

WALSH, J. B., 1984 How effective are selection and gene conversion in removing nonfunctional members of multigene families? Genetics **107** (Suppl): s112.

WALSH, J. B., 1985a Interaction of selection and biased gene conversion in a multigene family. Proc. Natl. Acad. Sci. USA **82:** 152–157.

WALSH, J. B., 1985b How many processed pseudogenes are accumulated in a gene family? Genetics **110:** 345–364.

WHITE, J. H., K. LUSNAK and S. FOGEL, 1985 Mismatch-specific post-meiotic segregation frequency in yeast suggests a heteroduplex recombination intermediate. Nature **315:** 350–352.

Communicating editor: B. S. WEIR

APPENDIX: CALCULATION OF $\pi(1)$ FOR THRESHOLD SELECTION

Our first order Taylor approximation leading to (6) is not valid for threshold fitness functions, so we proceed directly from (1b). Let the inflection point in the fitness curve occur at state k for a type I threshold (Figure 1A), and let it occur at $(n - k)$ for a type II threshold. This implies that for fixed s, both thresholds experience the same total amount of selection. We assume $|s| \ll 1$. For a type I threshold, we define the fitness of an individual that carries allele a at a fraction x of its loci, $w(x)$, as

$$w(x) = 1 + (x/k)sn \qquad 0 \le x \le k/n$$

$$w(x) = 1 + s \qquad k/n \le x \le 1,$$

which is plotted in Figure 1A. From (5), we have that for $i \neq k$ that $v_i/u_i = \exp\{-4N_e\xi^i\}$, where $\xi^i = \xi^0$ for $1 \le i < k$, and $\xi^i = \xi^1$ for $k < i \le n - 1$, where

$$\xi^0 = 2\gamma\delta + s/2k, \qquad \xi^1 = 2\gamma\delta. \tag{A.1}$$

For $i = k$, we use the above fitness function to compute v_i and u_i directly:

$$\frac{v_k}{u_k} = \frac{\xi^0}{\exp\{4N_e\xi^0\} - 1} \frac{1 - \exp\{-4N_e\xi^1\}}{\xi^1}. \tag{A.2}$$

Substitution of the above results into (1b) gives

$$\pi(1)^{-1} = 1 + \sum_{i=1}^{k-1} [r \exp\{-4N_e\xi^0\}]^i + \frac{v_k}{u_k} r^k \exp\{-4N_e(k-1)\xi^0\}\left(1 + \sum_{i=1}^{n-k-1} [r \exp\{-4N_e\xi^1\}]^i\right), \tag{A.3}$$

which we can reduce to

$$\pi(1)^{-1} = \frac{1 - [r \exp\{-4N_e\xi^0\}]^k}{1 - r \exp\{-4N_e\xi^0\}} + \Xi \frac{1 - [r \exp\{-4N_e\xi^1\}]^{n-k}}{1 - r \exp\{-4N_e\xi^1\}}, \tag{A.4}$$

where

$$\Xi = r^k \exp\{-4N_e(k-1)\xi^0\} \frac{\xi^0}{\exp\{4N_e\xi^0\} - 1} \frac{1 - \exp\{-4N_e\xi^1\}}{\xi^1} \tag{A.5}$$

For a type II threshold, $w(x)$ is given by

$$w(x) = 1 \qquad 0 \le x \le (1 - k/n)$$

$$w(x) = 1 + s(n/k)[x - 1 + k/n] \quad (1 - k/n) \le x \le 1,$$

which is plotted in Figure 1B. Proceeding in an analogous fashion as the above analysis, we find that we can express $\pi(1)$ for a type II threshold by using (A.4) and (A.5), provided we replace k by $(n - k)$ and set

$$\xi^0 = 2\gamma\delta, \qquad \xi^1 = 2\gamma\delta + s/2k. \tag{A.6}$$