Copyright © 1997 by the Genetics Society of America

The Evolution of Bacterial Transformation: Sex With Poor Relations

Rosemary J. Redfield,* Matthew R. Schrag† and Antony M. Dean‡

**Department of Zoology, University of British Columbia, Vancouver, Canada, V6T 1Z4, †Department of Mathematics, University of Chicago, Chicago, Illinois 60667 and ‡Department of Biological Chemistry, Chicago Medical School, North Chicago, Illinois 60064-3095*

Manuscript received October 17, 1996
Accepted for publication January 28, 1997

ABSTRACT

Bacteria are the only organisms known to actively take up DNA and recombine it into their genomes. While such natural transformation systems may provide many of the same benefits that sexual reproduction provides eukaryotes, there are important differences that critically alter the consequences, especially when recombination's main benefit is reducing the mutation load. Here, analytical and numerical methods are used to study the selection of transformation genes in populations undergoing deleterious mutation. Selection for transformability depends on the shape of the fitness function against mutation. If the fitness function is linear, then transformation would be selectively neutral were it not for the possibility that transforming cells may take up DNA that converts them into nontransformable cells. If the selection includes strong positive (synergistic) epistasis, then transformation can be advantageous in spite of this risk. The effect of low quality DNA (from selectively killed cells) on selection is then studied analytically and found to impose an additional cost. The limited data available for real bacterial populations suggest that the conditions necessary for the evolution of transformation are unlikely to be met, and thus that DNA uptake may have some function other than recombination of deleterious mutations.

CELLS of many bacterial species can actively take up DNA molecules from their environments (LORENZ and WACKERNAGEL 1994). Incoming homologous DNA frequently recombines with the cell's own genome and in so doing changes the cell's genotype. These natural transformation systems are commonly believed to be adaptations for genetic exchange (for competing viewpoints see REDFIELD 1993 and MICHOD and WOJCIECHOWSKI 1994). Although we are starting to understand the temporal regulation of the ability to take up DNA (competence) (GROSSMAN 1995; MACFADYEN *et al.* 1996), we remain ignorant about the evolutionary consequences of DNA uptake and transformation for bacteria.

The obvious comparison with meiotic sex in eukaryotic organisms is problematic for two reasons. First, unlike sexual recombination, bacterial transformation is infrequent, fragmentary, and nonreciprocal: most cells rarely take up DNA, and when they do, they replace only short segments of their genomes (MORRISON and GUILD 1973; GOODGAL 1982). Second, there is as yet no agreement on the function of meiotic sex itself. The leading contenders are elimination of deleterious mutations (KONDRASHOV 1994) and escape from rapidly evolving parasites and pathogens (HAMILTON *et al.* 1990; HOWARD and LIVELY 1994), although the importance of either has yet to be firmly established.

Corresponding author: Antony M. Dean, Department of Biological Chemistry, Chicago Medical School, 3333 Green Bay Rd., N. Chicago, IL 60064-3095. E-mail: deana@mis.finchcms.edu

REDFIELD (1988) modeled pure populations of transforming cells and found that (1) transformation could reduce mutation load and (2) the benefit was reduced or eliminated when the DNA came from cells killed by selection against deleterious mutations. However, no work has been reported on the dynamics of mixed populations of transformers and nontransformers, where transformers risk transforming themselves into nontransformers.

In this paper we develop an analytical model of transformation to investigate the assumption that the benefit of transformation comes from its ability to recombine deleterious mutations. We use this model to identify factors that limit the ability of a transforming population to invade a nontransforming population. A finding that the conditions required for invasion are biologically unrealistic would suggest that this benefit is unable to account for the evolutionary success of natural transformation systems.

THE MODEL

To develop a continuous time, overlapping generation model of the evolution of transformation in mixed populations, we will first describe pure populations of nontransforming and transforming cells, then a mixed population, and finally consider the composition of the DNA pool. The analysis of this model is presented in the RESULTS.

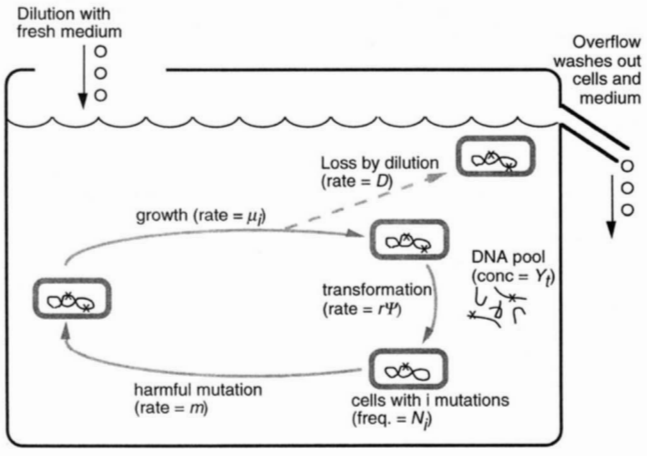

FIGURE 1.—Schematic of a chemostat population of cells undergoing transformation. Fresh medium enters the chemostat at a constant rate (D), while spent medium, cells and the extracellular DNA pool are washed out through an overflow at the same rate. Cells grow (rate = μ_i where i is the number of mutations in the genome) and are subject to mutation (rate = m) and transformation at a rate that is proportional to the concentration of extracellular DNA (rate = $r\Psi Y_t$). The extracellular DNA pool is generated by cell lysis.

Pure populations of nontransformers and transformers

We begin developing our model with the simplest cases, pure populations of nontransforming and transforming cells whose density is limited by dilution (as in a chemostat) and whose growth rates are limited by selection against deleterious mutations arising randomly at many loci. DNA fragments in the culture medium are taken up only by transforming cells. The basic outline of this model is presented in Figure 1.

A pure population of nontransformers: Deriving the equilibrium mutation distribution for a nontransforming population provides the foundation for subsequent analyses. Let N_i be the density of cells with i mutations in a chemostat population of nontransforming cells. Assume

$$\frac{dN_i}{dt} = (\mu_i - D - m)N_i + mN_{i-1}, \qquad (1)$$

where μ_i is the net growth rate (growth minus death; fitness) of a cell with i mutations, D is the dilution rate (the fractional rate of replacement of the mixed culture by fresh medium) and m is the mutation rate per genome per unit time.

When fitness is a decreasing function of i (*i.e.*, $\mu_i > \mu_{i+1}$), a mutation-selection balance will be attained where $dN_i/dt = 0$. Here, $dN_0/dt = (\mu_0 - D - m)N_0 = 0$ and thus $D = \mu_0 - m$. Summing Equation 1 over all i yields

$$\sum_{i=0}^{\infty} \mu_i N_i / N_t = D = \mu_0 - m, \qquad (2)$$

where $N_t = \Sigma_{i=0}^{\infty} N_i$. At this equilibrium the mean net growth rate of the population (its mean fitness) equals the dilution rate, and is lowered from the maximum growth rate (μ_0) by the mutation rate (m). Also,

$$N_i = N_{i-1}m/s_i,$$

where $s_i = \mu_0 - \mu_i$ is the selection coefficient. By induction

$$N_i = \frac{m^i}{\prod_{j=1}^{i} s_j} N_0 \quad i > 0, \qquad (3)$$

and the entire mutational distribution (N_i) is easily obtained.

A pure population of transformers: To describe a transforming population Equation 1 must be modified. Two additional terms in Ψ are added to describe how transformation adds cells to, and removes cells from, each mutational class.

$$\frac{dT_i}{dt} = (\mu_i - D - m)T_i + mT_{i-1} - \Psi Y_t T_i + \Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j. \qquad (4)$$

Ψ is the rate constant for transformation, Y_t is the total concentration of DNA in the environment (the DNA pool), and p_{ji} is the probability that a transformer with j mutations (T_j) is converted, by the DNA it takes up and recombines into its genome, into a cell with i mutations (calculation of p_{ji} is discussed below in the section on mixed populations). Thus, $\Psi Y_t T_i$ is the rate of loss of cells from this class by transformation and $\Psi Y_t \Sigma_{j=0}^{\infty} p_{ji}T_j$ is the rate of gain from the other classes. Note that when transformation does not change the net mutation number, cells removed by the first term in Ψ of Equation 4 are restored by the second.

A mixed population of transformers and nontransformers

To investigate the ability of transforming cells (T) to succeed in the presence of competing nontransformers (N) requires considering a mixed population. The difference between transformers and nontransformers is assumed to be genetically determined by one or more competence genes (c_1, c_2, etc.), with transformers carrying c^+ alleles and nontransformers carrying c^- alleles. Consequently, the DNA pool of a mixed population will contain both c^+ and c^- alleles, and transforming cells may sometimes take up and recombine c^- alleles, converting themselves into nontransforming cells. (We assume that c^+ and c^- alleles are not subject to the general

mutation process, and are themselves selectively neutral.)

Such a mixed population satisfies the following:

$$\frac{dN_i}{dt} = (\mu_i - D)N_i + mN_{i-1} - mN_i, \quad (5)$$

$$\frac{dT_i}{dt} = (\mu_i - D)T_i + mT_{i-1} - mT_i - \Psi Y_t T_i + \Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j - \left[\sum_{k=1}^{n} \rho_k\right]\Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j, \quad (6)$$

$$\frac{dC_{k.i}}{dt} = (\mu_i - D)C_{k.i} + mC_{k.i-1} - mC_{k.i} + \rho_k \Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j, \quad (7)$$

where ρ_k is the probability that a recombined fragment carries k c^- alleles and n is the number of competence genes per nontransformer (N_i) genome. Equation 5 is identical to Equation 1, while Equation 6 differs from Equation 4 only by an additional term in $\rho_k([\Sigma_{k=1}^{n} \rho_k]\Psi Y_t \Sigma_{j=0}^{\infty} p_{ji}T_j)$, which describes the net flux of T cells into the nontransforming C classes (Equation 7) through the recombination of DNA fragments harboring one or more c^- alleles. This "cost of conversion" is simply the net rate of transformation to T_i ($\Psi Y_t \Sigma_{j=0}^{\infty} p_{ji}T_j$) multiplied by the probability that a recombined fragment carries one or more c^- alleles ($\Sigma_{k=1}^{n} \rho_k$). We defer a discussion of ρ_k until later. Note that the transformers (T_i), the nontransformers (N_i), and the converts ($C_{k.i}$) have the same fitnesses (μ_i) and mutation rate (m), and so the success or failure of the transformers in competition arises from the ability of recombination to redistribute mutations and from the cost engendered by the risk of conversion.

How mutation number is changed by transformation

The effect of transformation is critically dependent on p_{ji}; the probability a cell with j mutations is converted by transformation into a cell with i mutations. This section describes the dependence of p_{ji} on both the number of mutations in the DNA fragment the cell takes up from the DNA pool, and the number of mutations in the chromosomal segment this fragment replaces.

Composition of the DNA pool: We defer consideration of the biological processes generating the DNA pool and for now assume only that cells produce DNA fragments at a rate θ that is proportional to cell density and independent of cell genotype. This allows derivation of the equilibrium distribution of mutations among these DNA fragments.

The density of fragments carrying j mutations changes according to

$$\frac{dY_j}{dt} = \theta \sum_{i=j}^{\infty} \binom{i}{j} r^j (1-r)^{i-j}\left(N_i + T_i + \sum_{k=1}^{n} C_{k.i}\right) - (D + E + \Psi T_t)Y_j, \quad (8)$$

where r is the size of DNA fragments expressed as a proportion of the genome. Thus, $\binom{i}{j} r^j(1-r)^{i-j}$ is the binomial probability that a donor cell with i mutations produces fragments with j mutations. The terms D, E and $\Psi T_t = \Psi \Sigma_{i=0}^{\infty} T_i$ describe loss of fragments by washout, by degradation, and by uptake by transforming cells. The DNA pool rapidly tends toward equilibrium values

$$Y_j = \frac{\theta \Sigma_{i=j}^{\infty} \binom{i}{j} r^j (1-r)^{i-j}(N_i + T_i + \Sigma_{k=1}^{n} C_{k.i})}{D + E + \Psi T_t}, \quad (9)$$

so long as $d(N_i + T_i + \Sigma_{k=1}^{n} C_{k.i})/dt \to 0$. Summing yields the total concentration of DNA fragments in the pool, $Y_t = \theta(N_t + T_t + \Sigma_{k=1}^{n} C_{k.t})/(D + E + \Psi T_t)$ where $C_{k.t} = \Sigma_{i=0}^{\infty} C_{k.i}$.

Effects of transformation: The probability, p_{ji}, that a cell with j mutations is transformed into a cell with i mutations is

$$p_{ji} = \sum_{k=0}^{\min(i,j)} \binom{j}{k} r^{j-k}(1-r)^k \frac{Y_{i-k}}{Y_t}. \quad (10)$$

Here, the coefficient $\binom{j}{k} r^{j-k}(1-r)^k$ is the binomial probability that a cell with j mutations takes up a DNA fragment differing by $j - i$ mutations from the genome segment it replaces, and Y_{i-k}/Y_t is the density of such fragments. For example, a cell harboring j mutations is converted into a cell with zero mutations with probability $r^j Y_0/Y_t$. This is simply the product of two probabilities: that the recombined DNA fragment harbors no mutations (Y_0/Y_t) and that it replaces a genomic fragment containing all the mutations in the cell (r^j). By extension, the probability that a cell harboring j mutations is converted into a cell with one mutation is simply $r^j Y_1/Y_t + jr^{j-1}(1-r)Y_0/Y_t$, where the first term is simply the probability that a recombined DNA fragment harboring one mutation replaces a genomic fragment containing all the mutations in the cell, and the second term is the probability that a recombined DNA fragment harboring no mutations replaces a genomic fragment containing all but one of the mutations in the genome.

The probability that a recombined fragment carries k c^- alleles is given by ρ_k, which depends not only upon the frequencies of N_t and $C_{k.t}$, but also upon the distribution of transformation loci across the genome and the size of DNA fragments (r) that are recombined. One might assume that the loci controlling transformation are randomly distributed in the genome, whence the probability that a recombined DNA fragment carries at least one c^- allele is $\Sigma_{k=1}^{n} \rho_k = [1 - (1-r)^n]$.

But this is inappropriate because these loci, few and certainly fixed in number and genomic position, constitute but one *sample* drawn from some (admittedly abstract) distribution. A unique sample does not constitute a distribution, so that $\Sigma_{k=1}^{n} \rho_k$ may differ wildly from naive expectation. Furthermore, different converts may have different constellations of k c^- alleles so that, in reality, the term $C_{k,t}$ describes a heterogeneous mixture of converts, each of which generates DNA fragments with 0, 1, 2 etc. c^- alleles at different probabilities. While tractable for defined cases, the mass of algebra generated by accounting for all such possibilities is prodigious.

Fortunately, the most biologically interesting case is where r, the size of the DNA fragments, is sufficiently small that the probability that a DNA fragment carries two or more c^- alleles is negligible. The competence loci are now unlinked, so that each convert carries exactly one c^- allele drawn from a sample of n c^- alleles in the nontransformers. The probability that a fragment of DNA carries a c^- allele is simply

$$\rho_1 = r(nN_t + C_{1,t})/K, \qquad (11)$$

where $K = N_t + T_t + C_{1,t}$ is the constant density of cells in the mixed population. The expressions for p_{ji} and ρ_1, when inserted into Equations 6 and 7, complete our mixed-population model.

RESULTS

The linear fitness function

Equation 3, the general expression for the equilibrium distribution of mutations in the nontransformers, is largely responsible for the complexity of Equations 6 and 7. However, a simple analytical solution of these equations is possible when fitness is a linear function of the number of mutations, *i.e.*, when $\mu_i = 1 - is$. This solution provides a benchmark against which the effects of other fitness functions can be compared.

A pure population of nontransformers: With a linear fitness function, the equilibrium density of cells with i mutations in an isolated population of nontransformers is (from Equation 3)

$$N_i = N_0 \frac{(m/s)^i}{i!},$$

and the mutational distribution at equilibrium is Poisson with mean m/s. The equilibrium distribution of mutations among DNA fragments in the environment is also Poisson, but with mean rm/s (compounding the binomial distributions of Equation 8 with the above Poisson distribution generates a Poisson distribution, see FELLER 1958).

A pure population of transformers: Suppose that, as with a pure population of nontransformers at equilibrium, the mutational distributions of the transformers and their DNA fragments are Poisson, with means m/s and rm/s respectively.
With

$$(\mu_i - D)N_i + mN_{i-1} - mN_i = 0$$

true for an equilibrium population of nontransformers, so

$$(\mu_i - D)T_i + mT_{i-1} - mT_i = 0$$

is true for this population of transformers. Furthermore, recombination has no effect on the mutational distribution of the transformers: the distribution in fragments from the DNA pool is identically Poisson to that of the genomic fragments they replace and thus net effect of transformation on T_i is

$$\Psi Y_t \sum_{j=0}^{\infty} p_{ji} T_j - \Psi Y_t T_i = 0$$

(see APPENDIX A). The model of a pure population of transformers reduces to

$$\frac{dT_i}{dt} = 0.$$

We conclude that should ever a pure population of transformers attain a Poisson distribution, mean m/s, then that distribution will be retained indefinitely because the growth rates are zero and independent of T_i and Y_i.

The question arises as to what happens if this distribution is perturbed, will it return to Poisson, or will it diverge to some other unknown distribution? In fact, it will return to the same Poisson distribution: that is, the Poisson distribution is Lyapunov asymptotically stable. To illustrate this, we conduct a local stability analysis by linearizing the model in the neighborhood of the Poisson equilibrium. Let

$$T_i = T_{i.\mathrm{eq}} + \delta T_i = T_t[(m/s)^i e^{-m/s}]/i! + \delta T_i$$

$$Y_i = Y_{i.\mathrm{eq}} + \delta Y_i = Y_t[(rm/s)^i e^{-rm/s}]/i! + \delta Y_i,$$

so that δT_i and δY_i represent small deviations from the Poisson equilibrium. These relations are substituted into the model described by Equations 4 and 8. Terms describing the Poisson equilibrium cancel and the equations are linearized by deleting the cross products because $\delta T_i \delta Y_j \to 0$. We now are left with a series of equations in $d\delta T_i/dt$ and $d\delta Y_i/dt$ that are linear in terms of δT_i and δY_i,

$$\begin{bmatrix} d\delta T_i/dt \\ d\delta Y_i/dt \end{bmatrix} = \mathbf{A} \begin{bmatrix} \delta T_i \\ \delta Y_i \end{bmatrix}.$$

Here, **A** is a matrix of constant coefficients drawn from Equations 4 and 8. Despite the fact that **A** is highly structured, we have yet to find an analytical solution.

However, numerical simulations using a wide range of values for mutation rates, selection coefficients, fragment sizes, rates of transformation, DNA production and degradation, demonstrate that the maximum eigenvalue $\lambda_{max} = 0$. This eigenvalue is associated with a positive eigenvector describing the equilibrium Poisson distributions of mutations in cells and in DNA fragments. Hence, this model of a pure population of transformers is uniformly stable and characterized by an asymptotically stable equilibrium distribution that is Poisson, mean m/s, when fitness is a linear function of the number of mutations ($\mu_i = 1 - is$).

These results are entirely in accord with those of REDFIELD (1988) who simulated pure populations of transformers and their DNA pools using a dynamic discrete generation time version of this model. Pure populations of transformers or nontransformers, initially devoid of mutations, each approach the same equilibrium mean fitness, a mean fitness that depends on the mutation rate, but not the selection coefficient, fragment sizes, and rates of transformation.

A mixed population of transformers and nontransformers: Again, suppose that the populations of transformers, nontransformers and converts are distributed as Poisson, with means m/s. Using the same approach as described in the above section, the mixed model (Equations 5–7) can be reduced to the following:

$$\frac{dN_i}{dt} = 0, \tag{12}$$

$$\frac{dT_i}{dt} = -\left[\sum_{k=1}^{n} \rho_k\right]\Psi Y_t T_i, \tag{13}$$

$$\frac{dC_{k.i}}{dt} = \rho_k \Psi Y_t T_i = \rho_k \Psi Y_t \left(\frac{T_t}{C_{k.t}}\right) C_{k.i}. \tag{14}$$

The transformers are now seen to be at a selective disadvantage, due entirely to the presence of c^- alleles in the DNA supply. The disadvantage is independent of both the mutation rate, the selection coefficient, and the dilution rate. By contrast, the intensity of selection is dependent on the values of ρ_k (the probability that recombined fragments carry k c^- alleles) and these are subject to change as converts accumulate in the population. Nevertheless, the Poisson distributions are retained as the transformers are purged from the mixed population, because the growth rates, $-[\Sigma_{k=1}^{n} \rho_k]\Psi Y_t$ and $\rho_k \Psi Y_t (T_t/C_{k.t})$, are independent of i.

Again, local stability analyses achieved by linearizing the model reveal that the largest eigenvalues for the nontransformers, the transformers, the converts and the DNA pool are each associated with Poisson distributions. Hence, the Poisson distributions represent a stable solution. This is hardly surprising because conversion events merely shuffle cells from one class to another and thus do not affect the distributions of deleterious mutations in those cells and the DNA pool.

When r, the size of the DNA fragments, is sufficiently small, the competence loci are unlinked. Each convert now carries exactly one c^- allele drawn from a sample of n c^- alleles in the nontransformers. The probability that a fragment of DNA carries a c^- allele is now given by Equation 11. With time, the mutational distributions converge on Poisson, and equations 12–14 become

$$\frac{dN}{dt} = 0, \tag{15}$$

$$\frac{dT}{dt} = -r\frac{nN + C_1}{K}\Psi Y_t T = -r\frac{nN + K - N - T}{K}\Psi Y_t T, \tag{16}$$

$$\frac{dC_1}{dt} = r\frac{nN + C_1}{K}\Psi Y_t T = r\frac{nN + C_1}{K}\Psi Y_t (K - N - C_1). \tag{17}$$

Integrating and taking the $\log_e[T/(N + C_1)]$ yields

$$\log_e [T_{t1}/(N_{t0} + C_{1.t1})] = \log_e [T_{t0}/K] - \log_e \left[1 - T_{t0}/K + \left(\frac{T_{t0}}{nN_{t0} + K - N_{t0}} - 1\right) \times (1 - e^{r\Psi Y_t (nN_{t0} + K - N_{t0})/K})\right], \tag{18}$$

where T_{t_1} and $C_{1.t_1}$, and N_{t_0} and T_{t_0} are the densities at time t_1 and t_0, respectively. The complexity of this function indicates that the intensity of selection changes with time. At $t = 0$, and taking $C_{1.t0} = 0$, the selection coefficient is simply $S = -rm\Psi Y_t$ (note that S refers to selection between the populations, whereas s refers to the fitness function). As time proceeds, the selection coefficient converges on $S = -r\Psi Y_t (nN_{t0} + K - N_{t0})/K = -r\Psi Y_t (nN_{t0} + C_{1.t\infty})/K$ as the converts replace the transformers. However, the selection is not inherently frequency dependent, and in the special case where there is only one c locus controlling transformation, $n = 1$ and Equation 18 collapses to

$$\log_e \left[\frac{T_{t1}}{N_{t0} + C_{1.t1}}\right] = \log_e \left[\frac{T_{t0}}{N_{t0} + C_{1.t0}}\right] - r\Psi Y_t t, \tag{19}$$

where $r\Psi Y_t$ is a conventional selection coefficient.

Equations 18 and 19 show that the selection is dependent only on the risk of being transformed by a c^- allele, that this may vary if $n > 1$, and that it is independent of the strength of selection (s) against the deleterious mutations in the linear fitness function ($\mu_i = 1 - is$).

We conclude that a linear fitness function yields an equilibrium distribution of deleterious mutations that

is Poisson, mean m/s, and that remains unchanged by recombination and conversion. Hence, recombination has no possible benefit, so that even a slight risk of conversion suffices to eliminate all c^+ alleles. To investigate conditions where c^+ alleles could be favored, we now consider nonlinear fitness functions.

Nonlinear fitness functions

The simple analytic solution above is not easily extended to the general case where fitness is an arbitrary decreasing function of the number of mutations per genome (*e.g.*, $\mu_i = 1 - (is)^\alpha$, $\alpha \neq 1$). The problem arises because an arbitrary fitness function generates nonlinear sets of differential equations that have no obvious solution.

Nevertheless, a solution can be obtained to the most biologically interesting case, invasion of an equilibrium population of nontransformers by a rare population of transformers. Under these circumstances the mutation distribution of the nontransformers remains similar to that of an isolated population, N_i^{iso}. Moreover, the DNA pool is dominated by DNA arising from the nontransformers. Consequently, the coefficients in Equation 6 behave as constants (since p_{ji} and ρ_k are now solely determined by N_i^{iso}), reducing a highly nonlinear system of simultaneous differential equations to a tractable linear form.

Selection when transformers are rare: The mixed model described by Equations 5–7 becomes

$$\frac{dN_i}{dt} = 0, \tag{20}$$

$$\frac{dT_i}{dt} = (\mu_i - D)\,T_i + mT_{i-1} - mT_i + \Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j - \Psi Y_t T_i - \left[\sum_{k=1}^{n} \rho_k\right]\Psi Y_t \sum_{j=0}^{\infty} p_{ji}T_j, \tag{21}$$

during invasion of an equilibrium population of nontransformers by a rare population of transformers. The converts, $C_{k,i}$, are ignored because they are so rare that they do not affect the DNA pool in any significant way.

Even this linearized model presents no obvious analytic solution. With $\alpha \neq 1$ the mutational distributions of the transformers, the nontransformers, and the DNA pool are no longer Poisson. Consequently, a simple solution, such as that derived for the linear fitness function with $\alpha = 1$, remains obscure. Nevertheless, matrix analysis can be used to characterize the selection coefficient acting on the transformers once they have reached an equilibrium distribution.

The DNA pool is determined by the equilibrium distribution of the nontransformers (N_i^{iso}) so that Equation 21 can be rewritten as

$$[dT_i/dt] = \mathbf{A}[T_i],$$

where **A** is a matrix of constant coefficients. We can apply a Lyapunov stability analysis of matrix **A** to determine whether a stable equilibrium distribution of T_i exists. With $(1 - \Sigma_{k=1}^{n} \rho_k) > 0$, all negative terms in Equation 21 are coefficients of T_i. Hence, **A** is irreducible and positive except on the main diagonal. Let $A = B + \lambda\mathbf{I}$, where **B** is an irreducible positive matrix. An irreducible positive matrix has exactly one nonnegative eigenvalue associated with a positive eigenvector (see GANTMACHER 1959). Consequently, **A** has at most one eigenvalue $\lambda_{max} \geq 0$ and this is associated with the equilibrium distribution of T_i.

The maximum eigenvalue of **A** is the selection coefficient of the transformers. Thus, if $S = \lambda_{max} > 0$ the invasion will be successful, and if $S = \lambda_{max} < 0$ the transformers will be purged. So long as the transformers remain very rare, all other eigenvalues are negative and the mutational distribution of T_i is Lyapunov asymptotically stable. In other words, given sufficient time and that T_i remains rare, the T_i will asymptotically approach an equilibrium distribution at which point λ_{max} provides an estimate of S, the selection coefficient of the transformers.

Simulating selection: In this section we investigate how the sign of $S = \lambda_{max}$ depends on the fitness function, fragment size, rate of transformation, and the mutation rate. First we set the stage by looking at the isolated populations of nontransformers (N_i^{iso}) that our transforming cells might invade. Figure 2 shows a series of fitness functions of the form $\mu_i = 1 - (is)^\alpha$, and the corresponding equilibrium distributions of mutations. For $\alpha = 1$ (solid line), the fitness function is linear and the mutation distribution is Poisson with mean $\kappa = m/s$ (here $\kappa = 30$). With positive epistasis ($\alpha > 1$), selection becomes relatively more intense as the number of mutations per cell increases. The mutational distribution shifts to the left, and κ is reduced. With negative epistasis ($\alpha < 1$), selection becomes relatively less intense as the number of mutations per cell increases. The mutational distribution shifts to the right, and κ rises. The shift in κ provides a measure of the effect of epistasis, and thus of the potential cost or advantage of recombination.

Next we use Equation 20 to evaluate the selection coefficient ($S = \lambda_{max}$) acting on transformers invading these equilibrium populations of nontransformers. The top panel in Figure 3 illustrates the effect of epistasis on the selection coefficient for transformation, using the fitness functions shown in Figure 2. Consistent with expectations from models of meiotic sex, transformation can be advantageous ($S > 0$) only with positive epistasis, while with negative epistasis it is always disadvantageous ($S < 0$). Yet even with positive epistasis, the advantage is decreased and even eliminated when

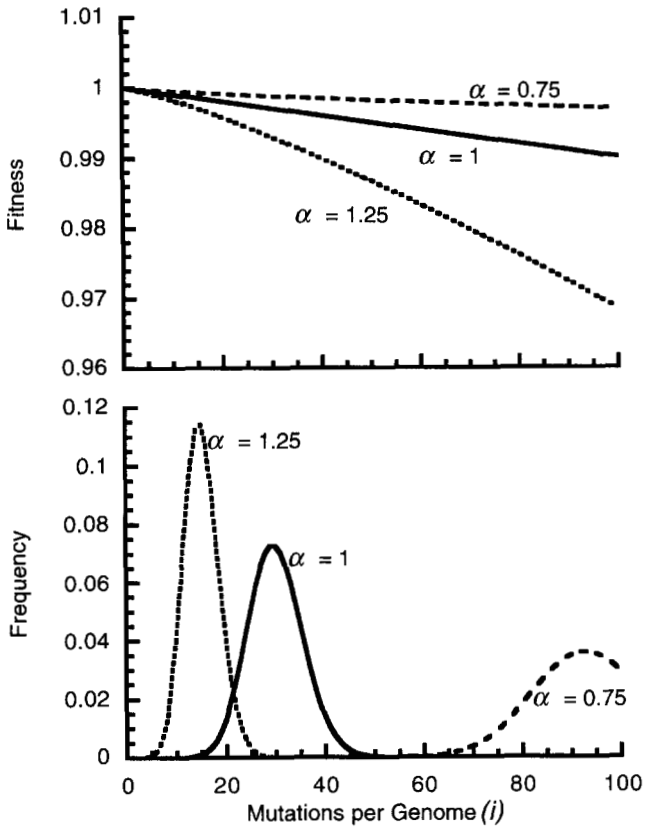

FIGURE 2.—Examples of the effect of epistasis on the fitness function ($\mu_i = 1 - (is)^{\alpha}$, top) on the equilibrium distribution of mutations among cells of an isolated population of nontransformers (N_i^{iso}, bottom); $m = 0.003$ (DRAKE 1991), $s_1 = s^{\alpha} = 10^{-4}$. In the absence of epistasis ($\alpha = 1$, —) the fitness function is linear ($\mu_i = 1 - is_1$), generating a Poisson distribution with mean $m/s_1 = 30$. With positive epistasis ($\alpha = 1.25$, ---) additional mutations cause a disproportional decrease in fitness and the mean number of mutations per cell is reduced. With negative epistasis ($\alpha = 0.75$, – – –) additional mutations have progressively less deleterious effects on fitness and the mean number of mutations per cell is increased.

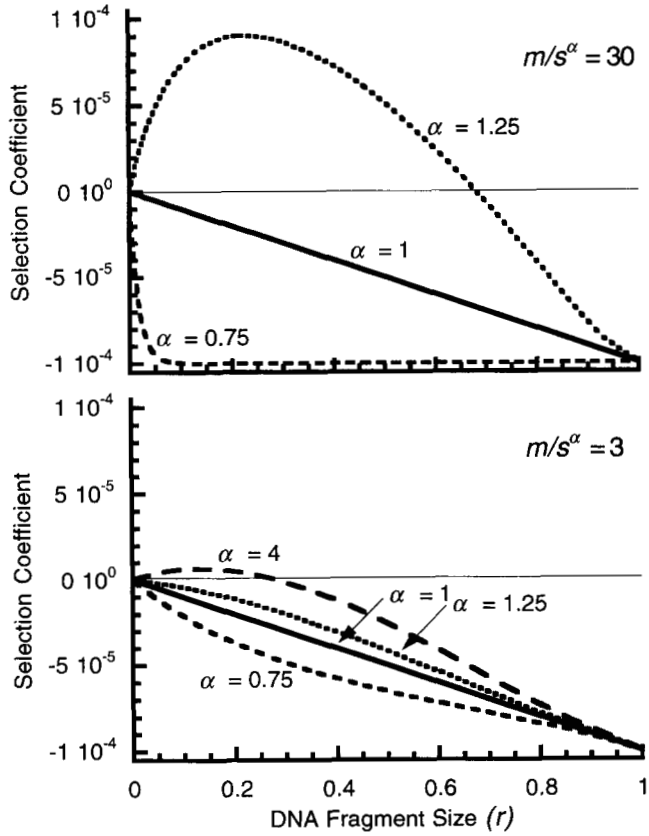

FIGURE 3.—The effect of recombined fragment sizes and selection intensities (s) on the selection coefficient for transformation (S). With $m/s^{\alpha} = 0.003/10^{-4} = 30$ and a modest degree of epistasis ($\alpha = 1.25$, ---), transformation is selectively advantageous (top panel) so long as recombined fragment sizes are not too large. When selection is intensified (lower panel; $m/s^{\alpha} = 0.003/10^{-3} = 3$), strong epistasis ($\alpha = 4$, – – –) is required if transformation is to be advantageous. In both cases transformation is infrequent ($\Psi Y_t = 10^{-4}$) and the number of transformation loci $n = 1$.

DNA fragments are large, because the risk of losing a c^+ allele overrides the weak selective advantage of recombining deleterious mutations. Note that as the number of loci controlling transformation (n) increases, so does the risk of losing a c^+ allele through recombination ($\Sigma_{k=1}^{n} \rho_k$).

The effect of intensifying the selection against deleterious mutations is to reduce the mean number of mutations per cell, as illustrated by the lower panel in Figure 3, where selection is increased from 0.0001 to 0.001. In the absence of epistasis this has reduced κ, the average number of mutations per nontransforming cell, from 30 to 3. Now, the transformers are only favored when epistasis is very strongly positive ($\alpha = 4$). This is because the mutation distribution remains approximately Poisson for moderate changes in epistasis when κ is small. Hence positive epistasis is necessary but not sufficient for transformation to be advantageous, and the stronger the selection against deleterious mutations, the weaker the selection for transformability.

Figure 4 illustrates the effect of changes in the net rate of transformation ΨY_t, under the strongest conditions from the top panel of Figure 3 ($m/s_1 = 30$, $\alpha = 1.25$). Low rates of transformation ($\Psi Y_t = 10^{-5}$) generate weak positive selection for transformers over a broad range of fragment sizes. High rates of transformation ($\Psi Y_t = 10^{-2}$) generate stronger positive selection but over a much narrower range of fragment sizes.

To summarize, the model suggests that the evolution of transformation requires positive epistasis ($\alpha > 0$), and that conditions are most favorable when the mean number of deleterious mutations per genome (κ) is large, when fragment sizes (r) are small, when the number of loci controlling transformation (n) is small, and when rates of transformation (ΨY_t) are moderate (if ΨY_t is large, then r must be tiny, and if ΨY_t is small, then selection may be too weak to override the effects of genetic drift).

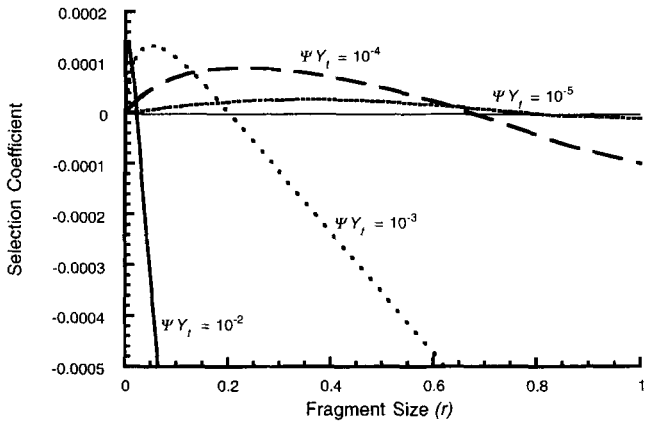

FIGURE 4.—The effect of transformation rates on the selection coefficient for transformation (S) with $m/s^{\alpha} = 0.003/10^{-4} = 30$, $\alpha = 1.25$ and $n = 1$. Increasing the rate of transformation (ΨY_t) causes selection intensities to increase at the expense of the range in recombined fragment sizes. Conversely, decreasing the rate of transformation causes selection intensities to decrease as the range in recombined fragment sizes broadens.

A simple approximation: There is a simple approximation (APPENDIX B) describing a necessary condition for the evolution of transformation. When fragment sizes are small ($r \rightarrow 0$),

$$S \approx r\Psi Y_t\left[\frac{m}{s_1} - \kappa - n\right]. \qquad (22)$$

Here $m/s_1 = m/s^{\alpha}$ is the value the Poisson mean number of mutations per genome would have if the fitness function were linear with $\mu_i = 1 - is_1$, κ is the actual mean number of mutations per nontransformer genome with the fitness function $\mu_i = 1 - (is)^{\alpha}$, and n is the number of unlinked loci controlling transformation.

The rate of recombination ($r\Psi Y_t$) influences the intensity of selection, but the direction is determined by terms in the brackets: for transformation to be selectively advantageous,

$$\frac{m}{s_1} > n + \kappa. \qquad (23)$$

This simple expression shows that selection in favor of the transformers may exist only when the Poisson mean exceeds the sum of the actual mean and the number of unlinked loci controlling transformation. Note that transformation is certain to be disadvantageous if either n or κ exceeds m/s_1.

The dependence on κ reflects the role of the epistatic interactions implicit in the fitness function. The inequality is not satisfied in the absence of epistasis ($\alpha = 1$ and $\kappa = m/s_1$), nor is it satisfied in the presence of negative epistasis ($\alpha < 1$ and $\kappa > m/s_1$). Only the presence of positive epistasis ($\alpha > 1$ and $\kappa < m/s_1$) potentially allows the evolution of transformation. The dependence on m/s_1 reflects the need for sufficient mutations to accumulate, and the dependence on n reflects the additional risk imposed by the chance of conversion at each locus required for transformation.

When the DNA pool contains excess mutations: To this point we have assumed that the rate of DNA production (θ) is proportional to cell density. One simple biological interpretation is that all cell deaths are random, with selection (*i.e.*, the fitness function) affecting only the rate of cell division. While many cell deaths are undoubtedly nonselective, many others may be caused by mutation. Consequently, it may be more valid to assume that the rate of cell division remains constant and mutations affect death rates. This does not affect the fitness function, nor the equilibrium distribution of mutations in the nontransformers, but it will change the composition of the DNA pool and thus the consequences of transformation. Specifically, selective deaths will cause the DNA pool to contain a disproportionately large number of mutations, thereby decreasing its potential usefulness.

Let the death rate be the selection coefficient, $(is)^{\alpha}$. The differential equation describing the rate of change of genomes, G_i, of dead cells in the DNA pool is

$$\frac{dG_i}{dt} = \theta'(is)^{\alpha}N_i^{iso} - (D + E + \Psi T_t)\,G_i$$

where θ', in analogy to θ for living cells, is the rate at which genomes from selective deaths ($(is)^{\alpha}N_i^{iso}$) enter the DNA pool. The equilibrium distribution of mutations among genomes from dead cells is calculated as

$$G_i = \frac{\theta'(is)^{\alpha}N_i^{iso}}{D + E + \Psi T_t} = \frac{\theta' mN_{i-1}^{iso}}{D + E + \Psi T_t},$$

because $(is)^{\alpha}N_i^{iso} = mN_{i-1}^{iso}$ (from Equation 3). The mean number of mutations per genome of dead cells is calculated by

$$\sum_{i=0}^{\infty} iG_i \Big/ \sum_{i=0}^{\infty} G_i = \sum_{i=0}^{\infty} \frac{i\theta' mN_{i-1}^{iso}}{D + E + \Psi T_t} \Big/ \sum_{i=0}^{\infty} \frac{\theta' mN_{i-1}^{iso}}{D + E + \Psi T_t}$$

$$= \kappa + 1.$$

Thus, the genomes of dead cells in the DNA pool contain, on average, one more mutation than those of the living cells, κ. This immediately suggests that the approximation in Equation 22 be modified to

$$S \approx r\Psi Y_t\left(\frac{\theta' m}{\theta}\right)\left[\frac{m}{s_1} - \kappa - n - 1\right], \qquad (24)$$

where the ratio $\theta' m/\theta$ normalizes the net rates of DNA production of the two hypotheses (a formal proof is given in APPENDIX C).

For transformation to be selectively advantageous

$$\frac{m}{s_1} > n + \kappa + 1. \qquad (25)$$

Transformers are selectively favored only when the Poisson mean exceeds the sum of the actual mean and the number of alleles controlling transformation plus one. Hence, the net effect of transforming with DNA arising from selective deaths is to raise the threshold by one mutation.

DISCUSSION

The above model was developed to identify factors constraining the evolution of genes causing bacterial transformation. The effect of transformation is to redistribute mutations among cells, and in so doing, influence the mutation load. In this respect, our model resembles those often applied to meiotic sexual reproduction, and it reveals the same dependence on the fitness function and mutation rate (CHARLESWORTH 1990; KONDRASHOV 1994). Our model also reveals that any potential benefit is also opposed by two novel costs that arise from the nonreciprocity of transformation: DNA derived from nontransforming cells carries nontransforming alleles (c^-) at the loci controlling transformability, and DNA derived from selectively killed cells carries excess deleterious mutations. If the role of transformation is to reduce mutational load, then both the fitness function and the transformation-specific costs constrain its evolution.

As with models of meiotic sex, transformation in our model is beneficial only when positive epistatic interactions between deleterious mutations exist (*i.e.*, $\alpha > 1$). The advantage is greater when the mutation rate per genome is high relative to the selective cost of a single mutation (that is, when m/s_1 is large), mainly because accumulated mutations are the raw material on which recombination acts. The potential benefit of this recombination is diminished by the risk of converting the alleles specifying transformability; maintaining each c^+ allele requires a corresponding increase in the strength of epistasis. Finally, transformation using DNA from selectively killed cells requires the effect of epistasis to be stronger still (by one additional mutation), because the excess mutations in the DNA pool prevent recombination from fully reducing the mutation load.

DNA is surely available for cells to take up. Many environmental and genetic factors cause cell lysis, including phage infection, colicins, antibacterial compounds produced by other organisms, and mutations affecting osmoregulation or the cell envelope. Thus some DNA will be available from random cell deaths, and some from selectively killed cells. If bacteria grow in colonies or dense populations, and if extracellular nuclease activity is low, conspecific DNA might even be abundant. Transformable bacteria can efficiently take up small DNA fragments from very dilute solutions (*e.g.*, BAROUKI and SMITH 1986), so even dilute or semidegraded DNA could be used.

Only a few genes are needed to confer transformability. The enzymes that carry out the physical recombination of DNA are present in all bacteria, because they are needed for DNA metabolism and repair (WALKER 1985; WALKER *et al.* 1985). Thus, any mutation that causes DNA fragments to be brought into a cell may lead to transformational recombination. In principle a single mutation, perhaps changing the substrate specificity of a permease, might be sufficient. The view that only a few genetic changes are needed to confer transformability is supported by the sporadic distribution of natural transformation systems among distantly related bacterial groups (LORENZ and WACKERNAGEL 1994).

That a transformation system be mechanistically and genetically accessible is insufficient for its evolution, it must also be favored by selection. A hypothetical population not subject to deleterious mutation has a fitness of 1, whereas a population of nontransformers subject to deleterious mutation has a fitness of $1 - m$ (Equation 2). Thus, the maximum conceivable gain in fitness from transformation is m. If $nr\Psi Y_t > m$, then the risk of conversion exceeds the benefit of transformation. Under these circumstances transformation can not evolve.

Experimentally measured mutation rates for a variety of microorganisms have been compiled by DRAKE (1991). Despite differences in genome size, the rates per genome all cluster around 0.003 per generation. Transformation experiments using genomic DNA and selecting for single alleles at unique loci mimic the introduction of a single c^- allele ($n = 1$) per generation. In laboratory experiments transforming bacteria provided with sufficient DNA usually replace ~1% of their genomes, (BARCAK *et al.* 1991; HOCH 1991). Hence if DNA is not very scarce, $r\Psi Y_t = 0.01 > 0.003 = m$. Note that the inequality becomes even greater with $n > 1$ ($rn\Psi Y_t > r\Psi Y_t$) and with a fair proportion of all mutations being selectively neutral: the experiments of KIBOTA and LYNCH (1996) suggest a lower bound of ~0.0002 for the genomic mutation rate to deleterious alleles in *Escherichia coli*. If these experimental values reflect those in natural populations, then transformation could never have evolved to reduce the mutation load.

Transformation rates are probably lower in natural populations than in laboratory experiments, where DNA is abundant and the cells are typically grown to maximum competence. Mutation rates in natural populations may also be higher. The possibility of $m > nr\Psi Y_t$ is therefore very real. Under these circumstances Equation 20, which uses DNA from selectively killed cells, provides a reasonable condition for transformation to be advantageous: with two loci controlling transformation, we need $m/s_1 - \kappa > 3$. A key requirement is that positive epistasis exists (*i.e.*, $\alpha > 1$), otherwise $m/s_1 - \kappa < 0$. Metabolic control theory suggests that combining

mutations in different genes within the same metabolic pathway will, in general, generate negative epistatic effects ($\alpha < 1$; REDFIELD 1988). The only experiment assessing fitness as a function of two metabolic fluxes provides no evidence that mutant genes in different pathways need generate epistatic effects ($\alpha = 0$; DYKHUIZEN and DEAN 1994). SZATHMÁRY (1993), analyzing a model of flux under stabilizing selection (DEAN *et al.* 1988; CLARKE 1991), concluded that positive epistasis ($\alpha > 1$) can be generated. Thus, experimental and theoretical results are, in the final analysis, equivocal with regard to the existence of positive epistatic effects.

Our model suggests that even when highly favorable conditions are assumed, the benefit derived from transformation reducing the mutational load is very slight (see Figure 3). For example, with $m = 0.03$ (high mutation rate), $\alpha = 3$ (very strong positive epistasis), $s_1 = 10^{-6}$ (weak selection), $\Psi Y_t = 0.01$ (moderate transformability because DNA will be degraded by exonucleases in natural environments), and that DNA comes from randomly rather than selectively killed cells, the selective advantage only reaches a maximum of $S \approx 0.01$ when $r = 0.065$. Yet a selective advantage of 0.01 is unlikely to overcome the physiological costs of the DNA uptake machinery. For example, selection against a constitutively expressed lactose operon (which encodes a permease) during growth on glucose incurs a cost of 0.05 (NOVICK and WEINER 1957).

Any analysis of bacterial transformation is severely limited by the lack of adequate estimates of natural mutation rates, and of the selective consequences of these mutations. Nevertheless, it seems likely that transformation has evolved for reasons other than reducing mutational load.

We thank two anonymous reviewers for their sterling efforts, their insights and their constructive criticisms, which all proved most helpful. We also thank RIK MYERS for suggesting a less salacious title. This work was supported by a grant to R.J.R. from the Natural Sciences and Engineering Research Council of Canada. R.J.R. is a Scholar of the Canadian Institute for Advanced Research's Evolutionary Biology Program.

LITERATURE CITED

BARCAK, G. J., M. S. CHANDLER, R. J. REDFIELD and J.-F. TOMB, 1991 Genetic systems in *Haemophilus influenzae*. Methods Enzymol. **204:** 321–342.

BAROUKI, R., and H. O. SMITH, 1986 Initial steps in *Haemophilus influenzae* transformation. Donor DNA binding in the *com10* mutant. J. Biol. Chem. **261:** 8617–8623.

CHARLESWORTH, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. **55:** 199–221.

CLARK, A. G., 1991 Mutation-selection balance and metabolic control theory. Genetics **129:** 909–923.

DEAN, A. M., D. E. DYKHUIZEN and D. L. HARTL, 1988 Theories of metabolic control in quantitative genetics, pp. 536–548 in *Proceedings of the Second International Symposium on Quantitative Genetics*, edited by B. S. WEIR, E. J. EISEN, M. M. GOODMAN and G. NAMKOONG. Sinauer Associates, Sunderland, MA.

DRAKE, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

DYKHUIZEN, D. E., and A. M. DEAN, 1994 Predicted fitness changes along an environmental gradient. Evol. Ecol. **8:** 1–18.

FELLER, W., 1958 *An Introduction to Probability Theory and its Applications. Vol I.* Wiley, New York.

GANTMACHER, F. R., 1959 *The Theory of Matricies. Vol. 2.* Chelsea, NY.

GOODGAL, S. H., 1982 DNA uptake in *Haemophilus* transformation. Ann. Rev. Genet. **16:** 169–192.

GROSSMAN, A. D., 1995 Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. Ann. Rev. Genet. **29:** 477–508.

HAMILTON, W. D., R. AXELROD and R. TANESE, 1990 Sex as an adaptation to resist parasites. Proc. Natl. Acad. Sci. USA **87:** 3566–3573.

HOCH, J. A., 1991 Genetic analysis in *Bacillus subtilis*. Methods Enzymol. **204:** 305–320.

HOWARD, R. S., and C. M. LIVELY, 1994 Parasitism, mutation accumulation and the maintenance of sex. Nature **367:** 554–557.

KIBOTA, T. T., and M. LYNCH, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature **381:** 694–696.

KONDRASHOV, A., 1994 Sex and deleterious mutation. Nature **369:** 99–100.

LORENZ, M. G., and W. WACKERNAGEL, 1994 Bacterial gene transfer by natural genetic transformation in the environment. Micro. Rev. **58:** 563–602.

MACFADYEN, L. P., I. R. DOROCICZ, J. REIZER, M. H. J. SAIER and R. J. REDFIELD, 1996 Regulation of competence development and sugar utilization in *Haemophilus influenzae Rd* by a phosphoenolpyruvate:fructose phosphotransferase system. Mol. Micro. **21:** 941–952.

MICHOD, R. E., and M. F. WOJCIECHOWSKI, 1994 DNA repair and the evolution of transformation IV. DNA damage increases transformation. J. Evol. Biol. **7:** 147–175.

MORRISON, D. A., and W. R. GUILD, 1973 Breakage prior to entry of donor DNA in *Pneumococcus* transformation. Biochim. Biophys. Acta **299:** 545–556.

NOVICK, A., and M. WEINER, 1957 Enzyme induction as an all or none phenomenon. Proc. Natl. Acad. Sci. USA **43:** 553–566.

REDFIELD, R. J., 1988 Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? Genetics **116:** 213–221.

REDFIELD, R. J., 1993 Genes for breakfast: the have your cake and eat it too of transformation. J. Hered. **84:** 400–404.

SZATHMÁRY, E., 1993 Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. Genetics **133:** 127–132.

WALKER, G. C., 1985 Inducible DNA repair systems. Ann. Rev. Biochem. **54:** 425–457.

WALKER, G. C., L. MARSH and L. A. DODSON, 1985 Genetic analyses of DNA repair: inference and extrapolation. Ann. Rev. Genet. **19:** 103–126.

Communicating editor: A. G. CLARK

APPENDIX A

Exchange by transformation: To show that the net rate of exchange by transformation with T_i is zero when the populations assume Poisson distributions (mean m/s), note that

$$\Psi Y_t \sum_{j=0}^{\infty} p_{ji} T_j - \Psi Y_t T_i = \Psi Y_t \sum_{j=0}^{\infty} (p_{ji} T_j - p_{ij} T_i).$$

Hence, it is sufficient to show that

$$T_j p_{ji} = T_i p_{ij}.$$

Starting with Equation 6 and substituting $T_j = T_t[(m/s)^j/j!]e^{-m/s}$ and $Y_{i-k} = Y_t[(rm/s)^{i-k}/(i-k)!]e^{-rm/s}$ yields

$$T_j p_{ji} = T_t \frac{(m/s)^j}{j!} e^{-m/s} \sum_{k=0}^{\min(i,j)} \binom{j}{k} r^{j-k} \times (1-r)^k \frac{(rm/s)^{i-k}}{(i-k)!} e^{-rm/s}$$

$$= T_t e^{-m/s} \sum_{k=0}^{\min(i,j)} \frac{r^{-k}(1-r)^k (rm/s)^{i+j-k}}{k!(j-k)!(i-k)!} e^{-rm/s}$$

$$= T_t \frac{(m/s)^i}{i!} e^{-m/s} \sum_{k=0}^{\min(i,j)} \binom{i}{k} r^{i-k}(1-r)^k \times \frac{(rm/s)^{j-k} e^{-rm/s}}{(j-k)!} = T_i p_{ij}.$$

Hence, the net rate of exchange by transformation with T_i is zero when the populations assume Poisson distributions (mean m/s).

APPENDIX B

An approximation: With $p_{j0} = r^j Y_0/Y_t$ and noting that $\mu_0 - D - m = 0$ because the nontransformers dominate and are distributed as if in isolation, rewrite Equation 20 with T_0 as

$$\frac{dT_0}{dt} = \left(1 - \left[\sum_{k=0}^{n} \rho_k\right]\right)\Psi Y_t\left[\sum_{j=0}^{\infty} r^j \frac{Y_0 T_j}{Y_t T_0} - 1\right] T_0 - \left[\sum_{k=1}^{n} \rho_k\right]\Psi Y_t T_0.$$

If the distribution of T_j is at equilibrium then the above immediately yields

$$S \approx \left(1 - \left[\sum_{k=1}^{n} \rho_k\right]\right)\Psi Y_t\left[\sum_{j=0}^{\infty} r^j \frac{Y_0 T_j}{Y_t T_0} - 1\right] - \left[\sum_{k=1}^{n} \rho_k\right]\Psi Y_t.$$

where Y_0 is calculated using Equation 9 with $N_j = N_j^{iso}$ (using Equation 3).

For $r \to 0$, the competence loci behave as if unlinked. Thus, the probability that a DNA fragment carries a c^- allele becomes $\Sigma_{k=1}^{n} \rho_k \approx nr$. Also, the mutational distribution of the transformers will remain similar to that of an equilibrium population of nontransformers in isolation. The selection coefficient for the transformers is then approximated by

$$S \approx -nr\Psi Y_t + (1-nr)\Psi Y_0\left[\sum_{j=0}^{\infty} r^j \frac{N_j}{N_0} - \frac{Y_t}{Y_0}\right].$$

Expanding the terms on the right hand side yields

$$S \approx -nr\Psi Y_t + (1-nr)\Psi Y_0\left[\left(\frac{r^0 N_0}{N_0} - \frac{Y_0}{Y_0}\right) + \left(\frac{r^1 N_1}{N_0} - \frac{Y_1}{Y_0}\right) + \left(\frac{r^2 N_2}{N_0} - \frac{Y_2}{Y_0}\right) + \cdots\right].$$

Noting that $(r^0 N_0/N_0 - Y_0/Y_0) = 0$ and substituting for N_i using Equation 3 yields

$$S \approx -nr\Psi Y_t + (1-nr)\Psi Y_0\left[\left(\frac{rm}{s^\alpha} - \frac{Y_1}{Y_0}\right) + \left(\frac{r^2 m^2}{s^\alpha (2s)^\alpha} - \frac{Y_2}{Y_0}\right) + \cdots\right],$$

which consists of an infinite sum of terms that decrease in a (roughly) geometric manner.

With $r \to 0$ terms in r^2 or greater powers will be very small and can be ignored. Furthermore the frequency of such small DNA fragments containing two or more mutations will be insignificant so that terms in Y_2 or greater can also be ignored. Thus, with $r \to 0$ the above reduces to

$$S \approx \Psi r Y_t\left[\frac{m}{s_1} - \kappa - n\right]$$

(Equation 22), because $Y_0 \approx Y_t$, $Y_1/Y_0 \approx n\kappa$, where κ is the actual mean number of mutations per genome for the arbitrary fitness function. $m/s_1 = m/s^\alpha$ is the value the Poisson mean number of mutations per genome would have if the fitness function were linear with $\mu_i = 1 - is_1$, and n is the number of alleles controlling transformation.

APPENDIX C

DNA from selective deaths: The density of DNA with j mutations arising as a consequence of selective death (Y_j') is given by Equation 9 with θ substituted by $\theta'(is)^\alpha$ so that the rate of DNA production is proportional to the selection coefficient:

$$Y_j' = \frac{\theta' \Sigma_{i=j}^{\infty} (is)^\alpha \binom{i}{j} r^j (1-r)^{i-j} N_i^{iso}}{D + E + \Psi T_t}.$$

From Equation 3, $N_i^{iso} = mN_{i-1}^{iso}/(is)^\alpha$ and

$$Y_j' = m\frac{\theta' \Sigma_{i=j}^{\infty} \binom{i}{j} r^j (1-r)^{i-j} N_{i-1}^{iso}}{D + E + \Psi T_t} = \theta' m \frac{r^j \Sigma_{i=j}^{\infty} \binom{i}{j} (1-r)^{i-j} N_{i-1}^{iso}}{D + E + \Psi T_t} - \frac{(1-r) r^j \Sigma_{i=j}^{\infty} \binom{i}{j} (1-r)^{i-j} N_i^{iso}}{D + E + \Psi T_t} + \frac{(1-r)}{\theta} Y_j\Bigg],$$

where Y_j is calculated using Equation 7. This equation can be reduced:

$$Y'_j = \frac{\theta' m}{\theta}\left[\frac{\begin{array}{l}\theta r \sum_{i=j}^{\infty} \binom{i}{j} r^{j-1}(1-r)^{i-j} N_{i-1}^{iso} \\ -(1-r)\binom{i}{j} r^{j-1}(1-r)^{i-j} N_i^{iso}\end{array}}{D+E+\Psi T_t} + (1-r)Y_j\right] = \frac{\theta' m}{\theta}$$

$$\times\left[\frac{\theta r \sum_{i=j-1}^{\infty} \binom{i}{j-1} r^{j-1}(1-r)^{i-(j-1)} N_i}{D+E+\Psi T_t} + (1-r)Y_j\right]$$

$$Y'_j = \frac{\theta' m}{\theta}\left[rY_{j-1} + (1-r)Y_j\right].$$

With

$$Y'_t = \sum_{j=0}^{\infty} Y'_j = \frac{\theta' m}{\theta}\left[rY_t + (1-r)Y_t\right] = \frac{\theta' m}{\theta} Y_t,$$

the selection coefficient, S', is approximated by

$$S' \approx -nr\Psi Y'_t + \Psi Y'_0\left[\left(\frac{rm}{s^\alpha} - \frac{Y'_1}{Y'_0}\right) + \left(\frac{r^2m^2}{s^\alpha(2s)^\alpha} - \frac{Y'_2}{Y'_0}\right) + \cdots\right],$$

which upon substituting Equation 19 becomes

$$S' \approx -nr\Psi \frac{\theta' m}{\theta} Y_t + \Psi \frac{\theta' \mathrm{m}}{\theta}(1-r)Y_0\left[\left(\frac{rm}{s^\alpha} - \frac{r}{(1-r)} - \frac{Y_1}{Y_0}\right) + \left(\frac{r^2m^2}{s^\alpha(2s)^\alpha} - \frac{rY_1}{(1-r)Y_0} - \frac{Y_2}{Y_0}\right) + \cdots\right]$$

$$\approx -nr\Psi \frac{\theta' m}{\theta} Y_t + \Psi \frac{\theta' m}{\theta}(1-r) \times Y_0\left[\left(\frac{rm}{s^\alpha} - \frac{r}{(1-r)} - \frac{(1+r)Y_1}{(1-r)Y_0}\right) + \left(\frac{r^2m^2}{s^\alpha(2s)^\alpha} - \frac{(1+r)Y_2}{(1-r)Y_0}\right) + \cdots\right].$$

When the leading term in the sequence is dominant,

$$S' \approx -nr\Psi \frac{\theta' m}{\theta} Y_t + \Psi \frac{\theta' m}{\theta}(1-r) \times Y_0\left(\frac{rm}{s^\alpha} - \frac{r}{(1-r)} - \frac{(1+r)Y_1}{(1-r)Y_0}\right).$$

The approximation reduces to

$$S' \approx r\Psi Y_t\left(\frac{\theta' m}{\theta}\right)\left[\frac{m}{s_1} - n - 1 - \kappa\right], \quad (24)$$

when r is small and $Y_0 \approx Y_t$, $Y_1/Y_0 \approx \pi\kappa$ and $m/s^\alpha = m/s_1$.