CHAPTER

8

Models for the Population Dynamics of Transposable Elements in Bacteria

Richard Condit

Transposable elements are segments of DNA capable of moving to new sites in the genome. They are widespread among both eucaryotes and procaryotes, probably occurring in all genomes. Besides being replicated during each generation as part of normal DNA synthesis, transposable elements (or transposons) also replicate during transposition, leaving a copy at the original site while inserting at a new site. This ability to transfer between genomes has led to concern about releasing genetically engineered organisms because of the possibility that introduced genes might be mobilized into transposons, onto plasmids, and spread into new organisms (Levy and Marshall 1988). So understanding the basic population dynamics of transposable elements—the processes that allow them to invade new bacterial strains and the factors that control their copy number in genomes—are important to biotechnology risk assessment. We should base decisions about the importance of gene transfer in genetically engineered microorganisms on a firm understanding of basic population processes.

Two alternative theories account for how transposons spread into new bacterial strains. According to one, transposons are parasitic DNA and can invade new populations despite the fact that they are detrimental to the

genome in which they reside (Doolittle and Sapienze 1980; Orgel and Crick 1980; Campbell 1981). The major alternative is that transposons improve the fitness of their host and so might be termed "mutualistic DNA." In bacteria, the latter hypothesis is particularly attractive because many transposons carry with them (during transposition) traits that are useful to the host bacterium, such as genes for antibiotic resistance or heavy metal resistance (Cohen 1976; Tanaka et al. 1983). Not all bacterial transposons carry beneficial genes, but there are other ways that they might benefit their host, such as by causing beneficial mutations (Reynolds et al. 1981; Chao et al. 1983).

One goal of population studies of transposable elements is to develop an experimental and theoretical system for testing the above hypotheses, and mathematical models are a major component of this system. The purpose of this paper is to review the use of mathematical models in studies of the population processes that account for the abundance, distribution, and movement of transposons in and between bacterial populations. There have been two modeling studies explicitly aimed at the study of bacterial transposons—one by Sawyer and Hartl (1986), and the other by myself, Frank Stewart, and Bruce Levin (Condit et al. 1988). I will not consider here the substantial literature on models for eucaryotic transposons (such as Hickey 1982; Charlesworth and Charlesworth 1983; Ginzburg et al. 1984).

Sawyer and Hartl were interested in the copy number of transposable elements in bacteria. Since transposons can replicate within a genome, copy number should build up through evolutionary time, until detrimental effects of the elements balance the increase. The Sawyer and Hartl model seeks to predict equilibrium copy number distributions based on various assumptions about transposition and its effect on fitness. Condit et al. (1988) were interested in a different issue—the invasion of a transposon into a new bacterial population due to its ability to transpose into new genomes.

Before continuing with a discussion of the models, it is necessary to provide a brief review of transposon biology and the terminology associated with it. In addition, discussion of the Condit et al. model requires some background in plasmid biology.

8.1 BACKGROUND AND TERMINOLOGY

In bacteria, transposable elements that carry no genes other than the two necessary for transposition are called "insertion sequences" and are given names that begin with "IS"—such as IS*4* and IS*10*. Elements that carry genes besides the basic two are called "Tn elements"—such as Tn*3*, Tn*5*, and Tn*10* (Campbell et al. 1977).

Most theoretical considerations of transposon biology assume that transposition is a *replicative* process, where transfer to a new site leads to an increase in copy number of the transposon. Indeed, where direct evidence is available, transposition is replicative (Klaer et al. 1980; Read et al. 1980).

It has been suggested, however, that some transposons are *conservative*, meaning that they move to a new site without leaving a copy behind (Berg 1977). From a population perspective, this is a very important distinction and one that must be considered in constructing a model. To avoid confusion, I define replicative and conservative strictly in terms of copy number—replicative transposition means a cell with one copy gets two, while conservative transposition leaves it with one—not in terms of molecular mechanisms. Transposons can also be *excised* from a genome, reducing by one the number of copies in the cell (Berg 1977; Bottstein and Kleckner 1977; Egner and Berg 1981; Iida et al. 1983).

In theory, rates of transposition (or excision) can be defined as the probability that an individual cell carrying a transposon will undergo transposition (or excision) per unit of time. Alternatively, this can be viewed as the proportion of cells in which the event occurs in the same time interval. In practice, accurate measurement of these rates can prove elusive.

Transposition is often *regulated* by the transposon itself. Multiple copies of an element within a cell tend to inhibit the transposition process, so that the rate of transposition may decline (Kleckner 1981).

Transposons are generally capable of moving between two sites within the same chromosome, or between two different pieces of DNA in the *same* cell, such as a plasmid and a chromosome, or two plasmids. Most transposons cannot move into *new* cells without the help of some other mechanism for DNA exchange. (A few transposons are able to move between cells, but they are exceptional and will not be considered here; see Gawron-Burke and Clewell 1982.) In eucaryotes, fusion of gametes from different individuals ("sex") provides such a mechanism, and this happens reliably once each generation. Gene transfer is less reliable and less well understood in procaryotes, occurring by one of three mechanisms—transduction, transformation, or conjugation. We chose to model the last one as the vehicle for transposon movement in our work (Condit et al. 1988).

Conjugation occurs when conjugative plasmids move between bacterial cells. Conjugative plasmids are small circles of DNA which inhabit bacteria, replicating and partitioning to daughter cells during cell division, and also capable of infecting new cells via conjugation (or mating). A cell with plasmids physically contacts one without, and the plasmid replicates and transfers between cells—two plasmid-bearing cells result. Plasmids can also be lost from individual cells through a process called *segregation*. Of particular significance for the present subject, plasmids often carry transposable elements.

8.2 MODELS

8.2.1 A Basic Model

As a simple illustration, which serves to introduce the structure of models for bacterial transposons, I first describe a very basic model: consider a set of interacting, time-dependent, dynamical systems that represent a mixture

of populations of bacterial cells, some without transposons and others with one or more elements. Let N_i be the population density of cells carrying i copies of a transposable element, and let Ψ_i be the growth rate of this i^{th} population. Transposition in the N_{i-1} population and excision in the N_{i+1} population create N_i cells, while either transposition or excision in the N_i population destroys N_i cells (the "N" symbols designate cell types as well as densities of those cell populations). Then the rate of change of each population (for $i>1$) can be written:

$$\dot{N}_i = \Psi_i N_i + \delta N_{i-1} - (\epsilon + \delta)N_i + \epsilon N_{i+1} \quad (1)$$

Here and in subsequent equations, a dot over a letter represents differentiation with respect to time. Transposition rate is represented by δ, and excision rate by ϵ.

Ψ_i is an important term because it defines any fitness effect of the transposable element. For example, if $\Psi_i<\Psi_0$ for all $i>0$, then the transposon is deleterious. In more familiar population genetics terminology, the selective coefficient s_i of the N_i population would be defined such that $\Psi_i = \Psi_0(1 - s_i)$.

The reason equation 1 cannot be applied when $i = 0$ and $i = 1$ is because cell type N_0 has no transposons and cannot be converted to type N_1 via transposition. Additional assumptions are needed to deal with the infection process; these will be treated below.

Both models considered here—that of Sawyer and Hartl and that of Condit et al.—are developed from this simple framework. But since they asked different questions, their analyses were quite different. Condit et al. examined this system when transposon-bearing populations are rare, and Sawyer and Hartl examined equilibria. In addition, as will become evident, there are other important differences between the models, particularly with regard to how the uninfected cells N_0 get their first transposon.

8.2.2 Sawyer and Hartl—A Model for Copy Number Distribution

Sawyer and Hartl (1986) were most interested in the dynamics of the copy number of a transposon in a population already carrying the element, and what regulates copy number. Copy number should be a result of the interplay between transposition, which drives copy number up, and excision and the fitness effect of the transposon, which tend to force it down. To simplify matters, however, Sawyer and Hartl ignored excision, since its rate is usually much lower than the transposition rate (Egner and Berg 1981; Foster et al. 1981b; Kleckner 1981).

8.2.2.1 Design of the Model Although their model is fundamentally very similar to the basic one described above, it differs in a few particulars. The growth rate of each population N_i is expressed as $\Psi - \Delta_i$, where Ψ is constant

and $\Delta_0 = 0$; hence, $\Psi - \Delta_i = \Psi_i$ of equation 1. The death rate caused by i transposons is given by Δ_i. Excision is ignored altogether. The problem of the infection of new cells is overcome in the simplest way possible: a single rate constant μ is defined as the infection rate.

This model leads to a series of differential equations:

$$\begin{aligned}\dot{N}_0 &= \Psi_0 N_0 - \mu N_0 \\ \dot{N}_1 &= \Psi_1 N_1 + \mu N_0 - \delta_1 N_1 - \Delta_1 N_1 \\ \dot{N}_i &= \Psi_i N_i + \delta_{i-1} N_{i-1} - \delta_i N_i - \Delta_i N_i.\end{aligned} \tag{2}$$

Note that absolute population size is not regulated in this model—it either grows or shrinks indefinitely. The only concern is with relative population sizes. Sawyer and Hartl also considered a model where population size is kept constant, in which the death of transposon-bearing cells is always accompanied by the creation of an equal number of transposon-free cells. This leads to only slightly different equilibria from the unregulated model given in equation 2.

The distribution of copy number of transposons is given by the relative populations of N_i. The equilibrium distribution will depend on transposition and death rates—δ_i and Δ_i—and how each changes with copy number i. If transposition is unregulated, δ increases linearly with i, and if each transposon copy causes the same decrement in fitness, then Δ increased linearly with i. On the other hand, if transposition were regulated, then δ might increase less than linearly, or even decrease with i.

Sawyer and Hartl's analysis consisted of finding equilibrium distributions of copy number for given examples of functions for δ_i and Δ_i. They found the equilibria by analyzing the system as a Markov chain, that is, each cell can be viewed as travelling a probabilistic pathway through different copy numbers. For example, a cell with one copy is converted to either one with zero, one, or two with defined probabilities, and then converted again, etc. The derivations of equilibria of Markov chains are beyond the scope of this chapter (see Sawyer and Hartl 1986).

An alternative procedure for finding equilibria would be to use the differential equations (see equation 2). N_{i+1}/N_i is constant at an equilibrium distribution of copy number, which is true when $\dot{N}_i N_{i+1} = \dot{N}_{i+1} N_i$. In general, this does not lead to explicit solutions for N_{i+1}/N_i, but the technique can be used for particular functions of Δ_i and δ_i. In the regulated model, where population sizes are constant at equilibrium, the distribution of copy number can be found by setting all derivatives to zero. This method can lead to explicit solutions for N_{i+1}/N_i.

8.2.2.2 Results. For the purpose of illustrating how the model works and some of its conclusions, I have calculated equilibrium copy number distributions for a couple of fitness and transposition functions used by Sawyer

et al. (1987), using the equations of the model with no population regulation as given in their paper:

1. Without any fitness detriment caused by the transposon ($\Delta = 0$), and without excision, copy number can only increase. Without any regulation of transposition rate, increase continues until all cells have an infinite number of copies, or until a copy number at which transposition ceases. This is a "null hypothesis" that spawned the modeling analysis. Clearly, in nature, something limits the copy number of genes that are capable of replicating themselves within the genome. There are other circumstances that lead to unlimited build-up of copy number, for example, if $\mu > \Delta_1 + \delta_1$ and $\Delta_i < \delta_i$.
2. The simplest model considered by Sawyer and Hartl is where δ and Δ are constant and independent of copy number. This represents strong regulation of transposition rate, since δ_i does not increase at all. The equilibrium distribution of copy number is geometric, with the ratio N_{i+1}/N_i a constant for all $i \geq 1$ (Figure 8–1). This is a rather nonintuitive result: since high copy number results in no greater mortality, and transposition creates more and more copies, one might expect copy number to build up indefinitely. The reason for a geometric decline in the abundance of cell types with higher copy number can be explained as follows: The density of any cell type N_i declines due to two forces—transposition to make N_{i+1} and death—but is created by only one event—transposition in the N_{i-1} population. Thus, for the relative sizes of N_i to be at equilibrium, N_{i-1} must be greater than N_i.

   The stationary distribution generated for any particular pair of functions δ_i and Δ_i and for the parameter μ depends only on their relative, not their absolute, values. This should be intuitive. All that matters is whether or not transposition rate is fast enough to overcome a fitness cost it engenders—the relative strength of the two forces. Higher parameter values will increase the speed with which the equilibrium is reached, though.
3. A simple alternative is when transposition and mortality rates are proportional to i: then $\Delta_i = i\Delta$ and $\delta_i = i\delta$ (Δ and δ are constants), and there is no regulation of transposition rate. In this situation, the ratio N_{i+1}/N_i is not constant at equilibrium, as in the model above, but declines to an asymptote as i increases. As illustrated in Figure 8–1, it is possible to generate a bimodal distribution of copy number with these functions, with modes at $i = 0$ and some $i > 0$.
4. Sawyer et al. (1987) considered four other functions for δ_i and Δ_i: harmonic, root, inverse root, and quadratic. Since each model required one function for each of the two variable parameters, a total of 36 models could have been analyzed (six functions for each parameter). Only nine were actually tested though, those that were most reasonable biologically.

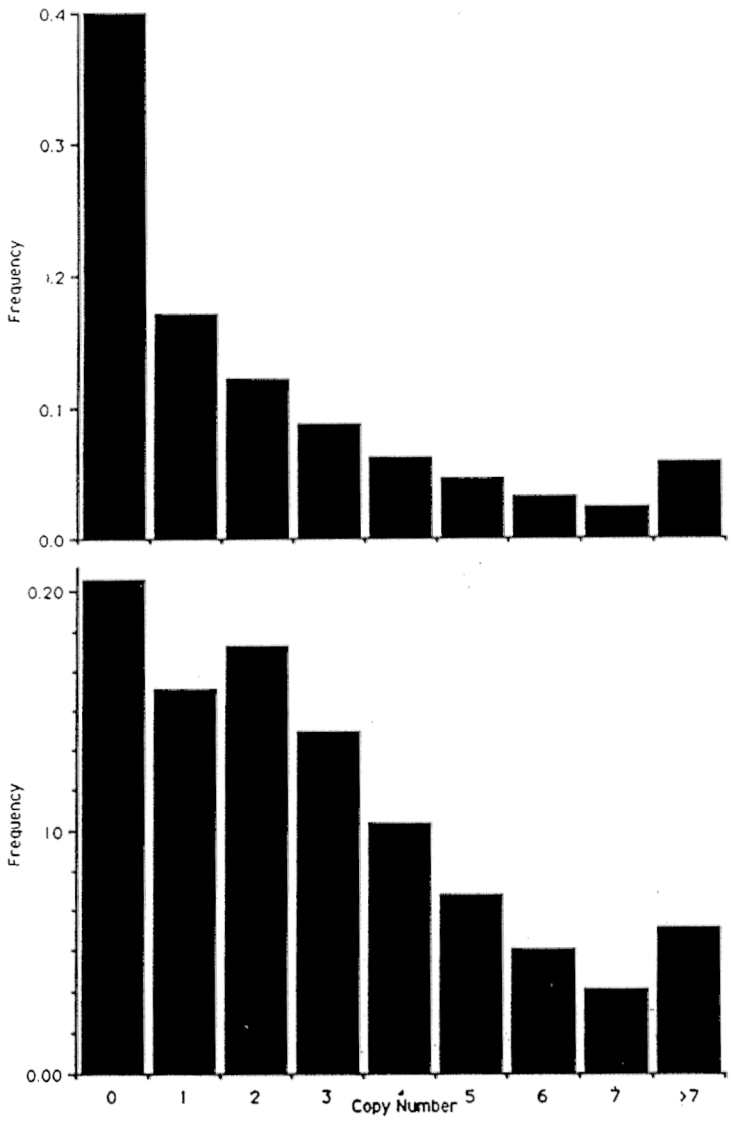

FIGURE 8–1 Theoretical distributions of transposon copy number at equilibrium, shown as the proportion of strains with various copy numbers. Top: constant model (transposition rate δ and mortality due to transposition Δ do not vary with copy number); $\delta = 10^{-6}$ and $\Delta = 10^{-4}$; infection rate = $\mu = 3 \times 10^{-5}$. Bottom: linear model; $\delta = 10^{-4}i$ and $\Delta = 6 \times 10^{-3}i$ (increasing linearly with copy number i); infection rate $\mu = 7 \times 10^{-5}$. These parameters were chosen in order to create a bimodal distribution. (The equation used to calculate the equilibrium distribution is given in Sawyer et al. 1987, p. 56; note that the variable symbols are different.)

Equilibrium distributions resulting from these models will not be given here.

8.2.2.3 Applications and Conclusions. One of the advantages of this modeling approach was that it was aimed at generating data of a sort that could be collected from natural populations. In fact, Sawyer and Hartl devised the model with the explicit intent of comparing theoretical copy number distributions with data collected on seven IS elements in 71 strains of *Escherichi coli* (Sawyer et al. 1987).

To fit models against the data, it was necessary first to estimate parameter values; the models allowed estimation of two parameters—μ/Δ and δ/Δ. A goodness-of-fit-test was then used to establish whether there was significant deviation between the distribution generated by any one model and the data for one IS element; with a total of nine models and seven IS elements, 63 tests were made.

Figure 8–2 shows the distribution of copy number for two insertion elements in the 71 strains. The distribution of IS*5* is typical. The largest copy number is zero, with declining numbers of strains carrying more and more elements. IS*3* is unusual, being the only element showing a bimodal distribution.

Unfortunately, no strong conclusions could be made from the model-fitting exercise. Nearly all of the models tested could be used to construct distributions that fit data from any of the insertion elements. The only exception was the bimodal IS*3*, for which only two models provided a reasonable fit. Since most models fit the data, it appears that the overall framework provides a useful description of the population dynamics of transposons. However, since all models fit, the approach could not distinguish the quantitative form of relationships between copy number and fitness or transposition rate.

However, one might suggest useful qualitative conclusions just by comparing the shapes of distributions from different elements. IS*4*, IS*5*, and IS*30* were quite rare—most strains had none of these elements, and very few had one or more. In contrast, IS*1* was common, with few strains uninfected and some strains with as many as 27 copies. It thus appears that regulation of copy number is much weaker in IS*1* than in IS*4*, IS*5*, or IS*30*. Either detrimental fitness effects are weaker for IS*1*, or there is less regulation of transposition rate as copy number increases.

In addition, one interesting feature of the distribution of insertion sequences in *E. coli* is that they never fit Poisson distributions, as *Drosophila* transposons do. Sawyer and Hartl (1986) argue that this is indicative of the lack of recombination between strains of *E. coli*.

8.2.2.4 Limitations of the Model. Two fundamental assumptions of this model need to be underscored in order to assess the generality of the conclusions:

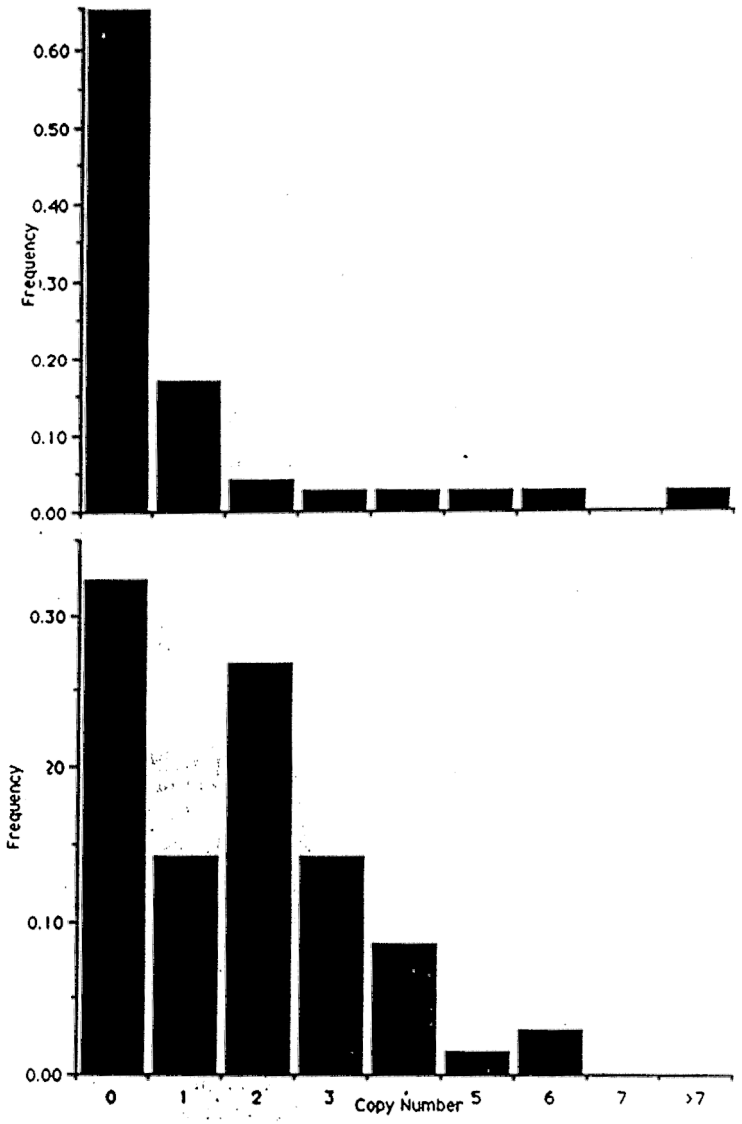

FIGURE 8–2 Actual copy number distributions for two insertion elements; frequency is the proportion of 71 strains with each copy number, as reported in Sawyer et al. (1987). Top: IS*5*. Bottom: IS*3*.

1. In any model-fitting exercise, one's conclusions are always limited by the scope of the models tested. Even if one model fits, one must consider that untested alternatives may have fit better. For example, transposons were assumed to be detrimental in terms of fitness in all the models examined by Sawyer et al. (1987). Could models for beneficial transposons (which Sawyer and Hartl did create but never tested) have provided better fits?
2. It was necessary to assume that all 71 strains examined were identical in parameter values in order to test the models. If strains differed in important ways, the statistical tests of goodness-of-fit would be invalid. The technique has no power to evaluate regulation of copy number *within* a single strain.

8.2.3 Condit et al.—A Model Describing Dynamics of Transposon Populations

The model by Condit et al. is based on principles similar to those described in the previous section. There is a mixture of populations with and without transposons, and cells change from one population to another by infection with a transposon, transposition, etc. But the goal of this model was to understand the dynamics of a population of transposable elements (within a population of bacterial cells), rather than equilibrium densities. For this reason, we wanted to develop a model that explicitly described the infection process and the dynamics of transposition.

Most transposons are unable to move between cells, so they can only infect new cells if there is some mechanism available for genome mixing. In bacteria, there are two reasonable candidates for providing this sort of "sexuality"—plasmids and phage. We chose plasmids in our model because plasmids commonly carry transposons in and between bacterial populations (Cohen 1976; Datta and Hughes 1983; Hawkey et al. 1985), and because models for their dynamics are available (Stewart and Levin 1977). Although phage *may* be important vehicles for transposons in natural populations, evidence that they are is lacking. The following model for plasmid dynamics is so intrinsic to the transposon model that it must be described in some detail.

8.2.3.1 The Plasmid Model. In the basic plasmid model developed by Stewart and Levin (1977), there are two cell populations, one without the plasmid and one with. The cell type designated N_0 has no plasmids, while type N_p does. Plasmid copy number is ignored—either a cell is infected, or it is not. Infection is assumed to happen at a rate proportional to the product of the two population densities, since plasmid transfer requires contact of N_0 and N_p cells. The rate constant of transfer is γ, defined as the probability per unit time that a single plasmid-free recipient will be infected if a single

plasmid-bearing donor is present. Also, assume cells lose plasmids at a constant rate τ. Then the dynamics of the system are described by the following equations:

$$\begin{aligned}\dot{N}_0 &= \Psi_0 N_0 - \gamma N_0 N_p + \tau N_p \\ \dot{N}_p &= \Psi_p N_p + \gamma N_0 N_p - \tau N_p\end{aligned} \tag{3}$$

The rate of growth of each population Ψ can be used to define a fitness cost of carrying the plasmid; define this as s, where $\Psi_p = \Psi_0(1 - s)$. If $s<0$, the plasmid confers a benefit. In addition, population regulation can be built in by making Ψ a function of a limiting resource concentration. Using chemostat models with saturation kinetics (Monod 1949) is a convenient way to model population regulation; readers should consult other works for details (Novick 1955; Dykhuizen and Hartl 1983).

A model for the dynamics of two interacting plasmids can easily be created by extending the basic model (Condit and Levin 1990): population N_1 carries plasmid 1, N_2 carries plasmid 2, and N_{12} carries both. Inclusion of the latter population is justified by reality—cells do routinely carry more than one plasmid type. A new conjugation parameter (γ_p) is needed to describe plasmid transfer into cells already carrying a plasmid, and a new segregation parameter (τ_p) for loss of one plasmid type from cells with two. Then:

$$\begin{aligned}
\dot{N}_0 &= \Psi_0 N_0 &&- \gamma N_0(N_1 + N_2 + N_{12}) + \tau(N_1 + N_2) \\
\dot{N}_1 &= \Psi_1 N_1 &&+ \gamma N_0 N_1 + \gamma N_0 N_{12}/2 - \tau N_1 - \gamma_p N_1(N_2 + N_{12}/2) + \tau_p N_{12}/2 \\
\dot{N}_2 &= \Psi_2 N_2 &&+ \gamma N_0 N_2 + \gamma N_0 N_{12}/2 - \tau N_2 - \gamma_p N_2(N_1 + N_{12}/2) + \tau_p N_{12}/2 \\
\dot{N}_{12} &= \Psi_{12} N_{12} &&+ \gamma_p N_{12}(N_1 + N_2)/2 + 2\gamma_p N_1 N_2 - \tau_p N_{12}
\end{aligned} \tag{4}$$

These equations are written in three parts. The first section to the right of the equal sign describes cell division; the second section describes dynamics between cells with plasmids and plasmid-free cells (identical to equation 3 above); the third section describes dynamics between cells with one plasmid and cells with two.

When cell type N_{12} acts as a donor, one of the two plasmids is transferred but not both; each is equally likely to be transferred (hence the terms $\gamma_p N_1 N_{12}$ and $\gamma_p N_2 N_{12}$ are divided by two). Segregation is assumed not to favor either plasmid, so that cells N_1 and N_2 are created equally often from N_{12}. In all our plasmid models, plasmid exchanges between identical cell types are not written into the equations because they do not cause cell transitions; however, one can demonstrate their occurrence experimentally and assume that they always occur.

As above, plasmid copy number is ignored. Cell type N_{12} has some of type 1 and some of type 2; segregation of one type means that all copies of that plasmid are lost. Neglecting copy number at this stage represents a loss of reality, since a cell with four copies of type 1 and one copy of type 2 would segregate quite differently from one with the opposite arrangement.

However, tallying copy number would add burdensome detail unnecessary for a description of the fundamental features.

Under many circumstances, it is reasonable to ignore plasmid-free cells, N_0. In experimental populations with plasmids present and persisting, cells without plasmids are usually not detectable, and probably have density below 10^{-3} of the plasmid-bearing population. This simplification will be used in the model of transposable elements. Later I will discuss the impact of plasmid-free cells, since in our previous work we did incorporate them in the model.

8.2.3.2 **The Transposon Model.** One of the fundamental differences between our model and the earlier work by Sawyer and Hartl is that we do not consider multiple copies of transposons on chromosomes or plasmids—a piece of DNA either has transposons or it does not. This simplification is acceptable because our goal was to describe the dynamics of infection, so movement of transposons within cells was not relevant. Sawyer et al. were interested in equilibrium copy numbers, so they had to consider intra-cell transposition and the build-up of copy number.

The transposon model of Condit et al. (1988) is essentially a model of two plasmids, as described above, where one of the plasmids carries the transposon and the other does not. The only complication is that transposition onto the chromosome must be included. Let plasmids without a transposon be number 1 while those with the element are number 2. Cells with a transposon on the chromosome are designated M, while those without are still N. Ignoring plasmid-free cells, the model requires six cell types: N_1, N_2, N_{12}, M_1, M_2, and M_{12}. All but cell type N_1 carry the transposon.

Let transposition occur at a constant rate δ per unit time in the population of cells carrying the element, and excision at rate ϵ. Assuming that transposition is replicative, then the cell transitions brought about by transposition are:

$N_2 \rightarrow M_2$ (transposon on plasmid moves to chromosome);
$N_{12} \rightarrow M_{12}$ (transposon on plasmid moves to chromosome);
$N_{12} \rightarrow N_2$ (transposon on plasmid 2 moves to other plasmid, turning it into plasmid 2);
$M_1 \rightarrow M_2$ (transposon on chromosome moves to plasmid); and
$M_{12} \rightarrow M_2$ (transposon on chromosome or on plasmid moves to other plasmid).

Other transposition events lead to a build-up in copy number of the element and are ignored.

Excision causes the following cell transitions:

$N_2 \rightarrow N_1$ (loss of transposon from plasmid);

$N_{12} \rightarrow N_1$ (loss of transposon from plasmid);
$M_1 \rightarrow N_1$ (loss of transposon from chromosome);
$M_{12} \rightarrow N_{12}$ (loss of transposon from chromosome);
$M_{12} \rightarrow M_1$ (loss of transposon from plasmid);
$M_2 \rightarrow N_2$ (loss of transposon from chromosome); and
$M_2 \rightarrow M_1$ (loss of transposon from plasmid).

In this model, it has been necessary to make assumptions about plasmid copy number: the above transitions are based on the assumption of only one copy of each plasmid per cell. Cell types N_{12} and M_{12} have one of both. It would be preferable to maintain generality and not consider copy number, but it does not seem feasible without adding considerable complexity.

If transposition is conservative, the cell transitions caused by transposition must be changed. The only relevant transfer occurs between plasmid and chromosome, causing interchange of cell types: $N_{12} \leftrightarrow M_1$ and $N_2 \leftrightarrow M_1$. In addition, transposition in M_2 acts just like an excision event, with either plasmid or chromosome losing its copy of the element.

To simplify writing the equations, define a population $P = (N_{12}/2 + M_{12}/2 + N_2 + M_2)$ as the density of all cells that are capable of donating the transposon-bearing plasmid (plasmid 2), and likewise, $Q = (N_{12}/2 + M_{12}/2 + N_1 + M_1)$ for cells donating plasmid 1. γ_1 and γ_2 are rate constants for transfer of plasmids 1 and 2, respectively. Otherwise, the assumptions are the same as those behind equation 4, and the resulting set of differential equations is:

$$
\begin{aligned}
\dot{N}_1 &= \Psi N_1 - \gamma_2 N_1 P + \tau N_{12}/2 + \epsilon(N_2 + N_{12} + M_1) \\
\dot{N}_{12} &= \Psi N_{12} + \gamma_1 N_2 Q + \gamma_2 N_1 P - \tau N_{12} + \epsilon M_{12} - 2\delta N_{12} - \epsilon N_{12} \\
\dot{N}_2 &= \Psi N_2 - \gamma_1 N_2 Q + \tau N_{12}/2 + \delta N_{12} + \epsilon M_2 - \epsilon N_2 - \delta N_2 \\
\dot{M}_1 &= \Psi M_1 - \gamma_2 M_1 P + \tau M_{12}/2 + \epsilon M_2 + \epsilon M_{12} - \delta M_1 - \epsilon M_1 \\
\dot{M}_{12} &= \Psi M_{12} + \gamma_2 M_1 P + \gamma_1 M_2 Q - \tau M_{12} + \delta N_{12} - 2\delta M_{12} - 2\epsilon M_{12} \\
\dot{M}_2 &= \Psi M_2 - \gamma_1 M_2 Q + \tau M_{12}/2 + \delta N_2 + \delta M_1 - 2\delta M_{12} - 2\epsilon M_2
\end{aligned}
\tag{5}
$$

Equations for an eight-cell model that includes plasmid-free cells can be found in Condit et al. (1988).

We have found two useful ways to analyze the equations. One is to use computer simulations based on Euler approximations, varying all parameters systematically to assess the impact of each on solutions to the equations. The other is to consider the situation where a transposon has just been introduced into a new population and is rare compared to the transposon-free cells, N_1. In this circumstance, the density of cell type N_1 can be treated as constant, and products of any two of the rare cell densities can be ignored. The set of equations is then linear and first order, and can be solved analytically, providing one uses a chemostat scenario where total cell density is approximately at equilibrium. Both approaches lead to the results discussed in the next section.

8.2.3.3 **Results of the Transposon Model.** We have been primarily interested in the fate of transposable elements that do not carry any useful function and so are either selectively neutral or deleterious. I return to the alternative scenario, beneficial transposons, below. All these results refer to chemostat models.

A replicative transposable element that is selectively neutral will invade a population of bacterial cells, providing there are conjugative plasmids present (Figure 8–3). In the absence of transposition ($\delta = 0$), otherwise identical conditions do not lead to invasion (see Figure 8–3), and if plasmid transfer is set to zero, no invasion occurs. The same result is obtained whether the plasmids are maintained by transfer or by benefitting their host, as long as some plasmid transfer occurs. When transposons are rare, their rate of invasion is constant per capita, so that the density of transposon-bearing cells increases exponentially (appearing linear on a semi-log plot, see Figure 8–3). We symbolize this invasion rate by I_0, where the "0" refers to the zero fitness effect of the element.

The rate at which a transposon invades turns out to be rather simple to predict, at least given the seeming complexity of the model. To formulate this prediction, it is useful to define an intuitively simple and important concept—"plasmid turnover" (symbolized by ρ), which is the total probability that any given cell in a population will exchange a plasmid per unit time. In a population without any transposon and a single plasmid type (cell N_1), the number of conjugation events per unit time (symbolized as C) is $\gamma_p(N_1)^2$, since all conjugation events happen between cells of the same type (these events were not included in equations 3, 4, and 5). Plasmid turnover is C/N_1 or $\gamma_p N_1$. This is the probability per unit time that a piece of plasmid DNA will find itself associated with a novel piece of chromosomal DNA, and likewise for chromosomes.

If transposition rate is much less than plasmid turnover, then the invasion rate I_0 will exactly equal the transposition rate. Conversely, if plasmid turnover is much less than the transposition rate, then I_0 will equal the plasmid turnover rate. That is

$$\begin{aligned} &\text{if } \delta \ll \rho, \text{ then } I_0 = \delta, \\ &\text{if } \rho \ll \delta, \text{ then } I_0 = \rho. \end{aligned} \tag{6}$$

Since ρ and δ were defined per unit time, invasion rate here takes the same units (time^{-1}). This is a simple and intuitively appealing result. Invasion of the transposable element requires that it infects new cells, which requires two steps—transposition and plasmid transfer. Like a chemical reaction, the rate of invasion is controlled by the slower of the two steps on which it depends.

The fact that two steps are necessary for invasion may not be immediately evident. At first consideration, it might appear that once a transposon is on a plasmid, then only plasmid transfer is necessary for invasion to occur, as if a nontransposing element could invade by hitch-hiking on a

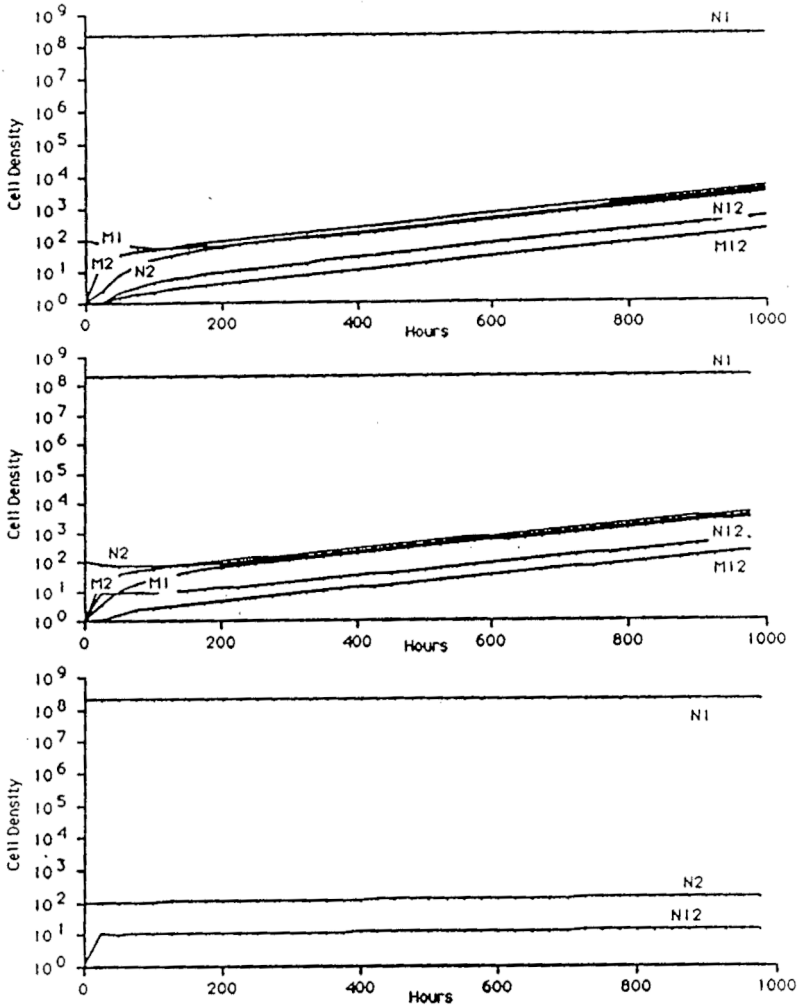

FIGURE 8–3 Simulations of changes in population density, based on equation 5 in the text. All populations maintained according to chemostat dynamics so that the total population reached an equilibrium of 1.9×10^8 cell ml^{-1}. Transposon-free cells (N_1) are at equilibrium prior to introduction of transposons at density of 10^2 cell ml^{-1}. In all three runs, $\gamma_1 = \gamma_2 = 10^{-10}$ ml cell^{-1} hr^{-1}; $\tau = 0.45$ hr^{-1}; $\epsilon = 0$ hr^{-1}; and there are no differences in fitness among the six cell types. Top: Transposon is introduced on the chromosome in cell type M_1, and transposition rate $\delta = 10^{-2}$. This value for δ is unrealistically high just to illustrate the invasion process; if δ were given a more realistic value (10^{-5} or lower), the increase would be visible only after about 10^6 hours (more than 100 years). Middle: Identical parameters, but the transposon is introduced on the plasmid in cell type N_2. Bottom: As above, but the transposition rate is set to zero.

plasmid. The reason this is wrong is that the plasmid population is assumed to be at equilibrium prior to the appearance of the transposon. Therefore, every plasmid transfer event, which adds one plasmid copy, must be matched by a segregation event, which eliminates one copy; if this were not true, then the number of plasmids would not be at equilibrium. If transposition never occurred, the transposon-bearing plasmid could not increase in abundance, since it must segregate as often as it transfers. But by transposing to the chromosome when arriving in a new cell, the element is protected against plasmid segregation. Thus transposition is necessary for the element to gain in frequency.

This conclusion, that invasion rate is controlled by the slower of two critical steps, provides a basic illustration of the functioning of the model—how a transposable element will behave in a model system of plasmids. Several other conclusions can be drawn by extending the analysis, and are justified further in Condit et al. (1988):

1. A conservative transposon that is selectively neutral or deleterious can never invade.
2. Even if a cell carrying a transposon suffers a fitness cost, it may still invade, providing the cost is not too high. Define the selective disadvantage as s, where $\Psi_t = \Psi_1(1 - s)$, Ψ_1 is the growth rate of the cell population without transposons, and Ψ_t is the growth rate for all cell types with the transposon. Then a simple conclusion can be drawn regarding the threshold level for s: if a transposon with no fitness cost has invasion rate I_0, then the same element will still invade with any fitness burden $s<I_0$. In fact,

$$I_s = I_0 - s. \tag{7}$$

   (Actually, it is necessary to match units in order to achieve equality. Invasion rate has been defined per unit time interval, and the s defined here has units of "per generation," so a correction for generation time is necessary; see Condit et al. 1988.)
3. If plasmid-free cells are added to the population, then all of the above conclusions hold, providing the term for plasmid turnover, ρ, is adjusted. When plasmid-free cells, N_0, are available, $\rho = [\gamma N_0 N_1 + \gamma_\rho (N_1)^2]/(N_0 + N_1)$. As above, this is the total number of conjugation events divided by the total number of cells.

In addition, I have repeated many of the analyses of transposon invasion using an alternative to the chemostat approach described by Condit et al. (1988). In the alternative, a small population of cells is provided with a large amount of resource at time zero, grows exponentially until the resource is depleted, and a small fraction of the cells are then transferred to fresh resource. This is known as serial transfer and is a common experimental tool for microbiologists (Atwood et al. 1951). Unlike a chemostat, bacterial

populations under serial transfer are not at equilibrium, which is not conducive to mathematical analysis of the differential equations. Nevertheless, when I repeat simulations in a serial transfer regime, results are qualitatively similar to those in a chemostat. Replicative transposons still invade, providing cost is not high, and conservative transposons do not. The exact equalities in equations 6 and 7 no longer hold, however.

As with Sawyer and Hartl's, our model was explicitly designed for direct testing with data from real populations. Laboratory populations of *E. coli*, the plasmid R100, and transposons Tn*3* and Tn*5* could be arranged to mimic the invasion-when-rare scenario illustrated in Figure 8-3 (Condit 1990). We found that the models quite successfully predicted changes in cell density caused by plasmid transfer, but because transposition rates were so low, we were unable to observe changes in frequency caused by transposition. That is, although the experimental data matched theoretical predictions, the tests of the model were not robust because transposition rate was never high enough to play a role in population dynamics.

8.2.3.4 Preliminary Models for the Dynamics of Beneficial Transposons. Initially, we felt that a situation where a transposon raises the fitness of its host was not particularly interesting. After all, any gene that raises host fitness will invade a population when rare, whether it transposes are not. If the fitness difference s (now s favors cells *with* a transposon) is much greater than I_0, then transposition would play a trivial role—invasion would be largely caused by differential fitness. On the other hand, if s is much lower than I_0, then the transposon is essentially neutral, and we return to the conclusions above.

Recently, however, we have reconsidered scenarios with beneficial transposons and decided they are not so trivial, providing we change the structure of the model. Much of what follows is based on the model presented in Condit and Levin (1990) whose purpose was different, but the structure of this model has relevance for modeling transposable elements.

The problem with my initial thinking about beneficial transposons was that it considered only whether genes within the transposable element were beneficial, not how the *act of transposition* itself could have positive fitness. If a transposon carries useful genes, then of course one anticipates its spread into a population, but once fixed, the ability to transpose should be lost while the genes within the element are maintained.

What is necessary for transposability to be maintained is a situation where new gene combinations have a selective advantage. We modeled one such scenario: two plasmids co-occur in a population of bacteria, each carrying a gene that can be beneficial to the bacteria; both of the genes are contained within transposable elements. In addition, the two plasmids are "incompatible," that is, single bacterial cells cannot maintain both for long (Novick 1987). This is modeled by having a high segregation rate, τ_p, among

the cells N_{12}. If selection for both genes occurs simultaneously, so that cell type N_{12} has a selective advantage over N_1 and N_2, then a population of cells in which one of the transposons has moved between plasmids (or onto the chromosome) will invade. The reason is the instability of the plasmids. Cell type N_{12} will have the highest growth rate, but it will also lose plasmids at a high rate, creating progeny with low fitness. Cells with both transposons (and hence both useful genes) on the same plasmid do not suffer from this instability. We proved using simulations that the invasion rate of the new cell type is equal to the segregation rate of the two plasmids.

This situation leads to a selective advantage for *transposability* based on the unstable inheritance of separate pieces of DNA in the same cell. However, once the transposons find themselves in a new arrangement, the selective advantage to transposition is lost. The benefit to the transposon has only been temporary, unless new circumstances constantly adjust fitnesses and continually favor gene rearrangements. The problem of maintaining transposability thus appears to be the same as the general problem of the evolution of recombination and sex (Feldman et al. 1980). Clearly, further theoretical work is needed in this area.

8.2.3.5 **Limitations of the Models Describing Dynamics.** Some of the assumptions underlying the model have been mentioned already, but certain ones deserve particular attention before assessing the theoretical conclusions:

1. Populations were assumed to be at equilibrium before the transposon was introduced. If a plasmid carrying a transposon were introduced into a population formerly without that plasmid, the transposon could be swept to fixation by plasmid transfer alone. Perhaps nonequilibrium systems, with plasmids frequently invading and going extinct, offer a better avenue for invasion of selectively neutral transposons. In Condit et al. (1988), we argue that transposons are no more likely to invade nonequilibrium populations than equilibrium populations, but the idea has not been tested rigorously.
2. We did not consider whether transposons affect plasmids in any way. If a transposon is deleterious to its host, but somehow benefits the plasmid it is on (relative to other plasmids), its ability to invade would improve, but we have not quantitatively assessed this situation.
3. Our models explicitly model populations in liquid culture. Although I do not anticipate that surface populations would exhibit dramatically different behavior, the possibility should be considered further.

8.3 CONCLUSIONS

8.3.1 Explanations for the Invasion and Maintenance of Transposons

The models of Condit et al. (1988) and Condit and Levin (1990) represent standard population genetic approaches. We sought to understand the basic dynamics of genes that are able to transpose between pieces of DNA in simple model systems. Obviously, the model systems do not reflect reality exactly, but it is necessary to understand dynamics in simple systems before considering more complicated real populations. What do the models show?

It is plausible for transposons to be "parasitic DNA." This had been taken as a given in bacterial populations (Campbell 1981), but never proven. However, as Condit et al. (1988) argue, there are several reasons to doubt whether the parasitic DNA hypothesis can be generally applicable to bacterial transposons. The problem is twofold: first, bacteria are almost asexual, since the rate of plasmid turnover is usually quite low (Levin et al. 1979; Freter et al. 1983); second, transposition rate is extremely low for nearly all bacterial transposons (Foster et al. 1981a; Peterson et al. 1982; Iida et al. 1983; Meyer et al. 1983; Schmidt and Klopfer-Kaul 1984). Both rates are critical for invasion as a parasite, and since both are so low, it must be a rare event for a transposon to establish itself in a new bacterial strain without benefitting its host.

The above statements about magnitudes of transposition and plasmid turnover are based on very little data (at least for natural populations), and they should be considered provisional. Also, one should consider transduction as an alternative gene transfer mechanism, so our conclusion in the previous paragraph should be debated. Still, it does not seem plausible that transposons have become so widespread as parasitic DNA.

In contrast, Sawyer et al. (1987) concluded that their models suggest "... a moderate to strong detrimental effect of copy number [of transposable elements] on fitness...." for most insertion elements. But their group did not examine models for beneficial transposons, and as they readily acknowledge, their conclusions were not strong. Clearly, more data about the dynamics of transposable elements in natural bacterial populations are needed to establish whether some are acting as parasitic DNA.

Models that demonstrate how transposition can be selectively advantageous are available (Condit and Levin 1990), but they do not provide a compelling general argument for the maintenance of transposability. Although transposition can have a transient advantage, the theory ends there. How can the ability to move to new sites be continuously selected for? More theoretical work is needed in this area.

8.3.2 Applications of Models to Risk Assessment

On the basis of our study of the dynamics of the spread of transposons, and information about how low the critical rate parameters are, we predict that engineered genes released into bacterial communities will never become

established in novel populations *as long as* they do not raise fitness of any strain they inhabit. Genes *will* transfer into new cells, but without a selective advantage these transfer events will be rare and isolated and will not lead to fixation of the gene at a new site.

Nevertheless, our models for beneficial transposons show that if a transposition event creates a new cell that has higher fitness than its progenitors, then a new strain carrying a rearrangement can become fixed very rapidly. This raises an important caution for risk assessment. No matter how rare genetic rearrangements might be, they can readily lead to new strains if the selective conditions are right.

It is debatable whether general predictions like this are useful for risk assessment. One should probably make assessments on a case-by-case basis, and it is possible to apply models to individual cases. In Condit (1988), I describe how mathematical models of bacterial populations might be used to make predictions about the fate of engineered bacteria released into the environment. Our models for the dynamics of transposon populations could be used to make such predictions. Given parameters for transposition, plasmid transfer, and selective coefficients, one could try to predict the fate of a particular engineered gene released into a particular bacterial population.

Although this might be a goal for modeling work in risk assessment, quantitative predictions based on detailed models are unlikely to be accurate enough as a basis for important decisions, at least in the near future. Instead, I see models as continuing to be a basic tool for evaluating the processes that control the abundance and distribution of transposable elements in natural populations of bacteria. This basic knowledge is crucial for making decisions about manipulating bacterial populations.

REFERENCES

Atwood, K.C., Schneider, L.K., and Ryan, F.J. (1951) *Cold Spring Harbor Symposia on Quantitative Biology* 16, 345–355.

Berg, D.E. (1977) in *DNA Insertion Elements, Plasmids, and Episomes* (Bukhari, A.I., Shapiro, J.A., and Adhya, S.L., eds.), pp. 185–203, Cold Spring Harbor Laboratory, New York.

Bottstein, D., and Kleckner, N. (1977) in *DNA Insertion Elements, Plasmids, and Episomes* (Bukhari, A.I., Shapiro, J.A., and Adhya, S.L., eds.), pp. 185–203, Cold Spring Harbor Laboratory, New York.

Campbell, A. (1981) *Ann. Rev. Microbiol.* 35, 55–83.

Campbell, A., Berg, D., Botstein, D., et al. (1977) in *DNA Insertion Elements, Plasmids, and Episomes* (Bukhari, A.I., Shapiro, J.A., and Adhya, S.L., eds.), pp. 15–22, Cold Spring Harbor Laboratory, New York.

Chao, L., Vargas, C., Spear, B.B., and Cox, E.C. (1983) *Nature* 303, 633–635.

Charlesworth, B., and Charlesworth, D. (1983) *Genet. Res. Camb.* 42, 1–27.

Cohen, S.N. (1976) *Nature* 263, 731–738.

Condit, R. (1988) in *Biotechnology: A Risk Assessment Framework* (Mittleman, A., ed.), pp. 88–157, prepared for the U.S. Environmental Protection Agency, Washington, D.C., by Technical Resources Inc., Washington, D.C.
Condit, R. (1990) *Evolution* 44, 347–359.
Condit, R., and Levin, B.R. (1990) *Amer. Natur.* 135, 573–596.
Condit, R., Stewart, F.M., and Levin, B.R. (1988) *Amer. Natur.* 132, 129–147.
Datta, N., and Hughes, V.N. (1983) *Nature* 306, 616–617.
Dykhuizen, D.E., and Hartl, D.L. (1983) *Microbiol. Rev.* 47, 150–168.
Doolittle, W.F., and Sapienza, C. (1980) *Nature* 284, 601–603.
Egner, C., and Berg, D.E. (1981) *Proc. Nat. Acad. Sci.* 78, 459–463.
Feldman, M.W., Christiansen, F.B., and Brooks, L.D. (1980) *Proc Nat. Acad. Sci.* 77, 4838–4841.
Foster, T.J., Davis, M.A., Roberts, D.E., Takeshita, K., and Kleckner, N. (1981a) *Cell* 23, 201–213.
Foster, T.J., Lundblad, V., Hanley-Way, S., Halling, S.M., and Kleckner, N. (1981b) *Cell* 23, 215–227.
Freter, R., Freter, R.F., and Brickner, H. (1983) *Infect. Immun.* 39, 60–84.
Gawron-Burke, C., and Clewell, D.B. (1982) *Nature* 300, 1–3.
Ginzburg, L.R., Bingham, P.M., and Yoo, S. (1984) *Genetics* 107, 331–341.
Hawkey, P.M., Bennett, P.M., and Hawkey, C.A. (1985) *J. Gen. Microbiol.* 131, 927–933.
Hickey, D.A. (1982) *Genetics* 101, 519–531.
Iida, S., Meyer, J., and Arber, W. (1983) in *Mobile Genetic Elements* (Shapiro, J.A., eds.), pp. 159–221, Academic Press, New York.
Klaer, R., Pfeifer, D., and Starlinger, P. (1980) *Molec. Gen. Genet.* 178, 281–284.
Kleckner, N. (1981) *Ann. Rev. Genet.* 15, 341–404.
Levin, B.R., Stewart, F.M., and Rice, V.A. (1979) *Plasmid*, 2, 247–260.
Levy, S.B., and Marshall, B.M. (1988) in *The Release of Genetically Engineered Micro-organisms* (Sussman, M., Collins, C.H., Skinner, F.A., and Stewart-Tull, D.E., eds.), pp. 61–76, Academic Press, London.
Meyer, J.F., Nies, B.A., and Wiedemann, B. (1983) *J. Bact.* 155, 755–760.
Monod, J. (1949) *Ann. Rev. Microbiol.* 2, 371–394.
Novick, A. (1955) *Ann. Rev. Microbiol.* 9, 97–110.
Novick, R.P. (1987) *Microbiol. Rev.* 51, 381–395.
Orgel, L.E., and Crick, F.H. (1980) *Nature* 284, 604–607.
Peterson, B.C., Hashimoto, H., and Rownd, R.H. (1982) *J. Bact.* 151, 1086–1094.
Read, H., Das Sarma, S., and Jaskunas, S.R. (1980) *Proc. Nat. Acad. Sci.* 77, 2514–2518.
Reynolds, A.E., Felton, J., and Wright, A. (1981) *Nature* 293, 625–629.
Sawyer, S., and Hartl, D. (1986) *Theor. Pop. Biol.* 30, 1–16.
Sawyer, S.A., Dykhuisen, D.E., Dubose, R.F., et al. (1987) *Genetics* 115, 51–63.
Schmidt, F., and Klopfer-Kaul, I. (1984) *Mol. Gen. Genet.* 197, 109–119.
Stewart, F.M., and Levin, B.R. (1977) *Genetics* 87, 209–228.
Tanaka, M., Yakamoto, T., and Sawai, T. (1983) *J. Bact.* 153, 1432–1438.