Role of Cryptic Genes in Microbial Evolution[1]

Barry G. Hall
University of Connecticut

Shozo Yokoyama
Washington University School of Medicine

David H. Calhoun
Mount Sinai School of Medicine

Cryptic genes are phenotypically silent DNA sequences, not normally expressed during the life cycle of an individual. They may, however, be activated in a few individuals of a large population by mutation, recombination, insertion elements, or other genetic mechanisms. A consideration of the microbial literature concerning biochemical evolution, physiology, and taxonomy provides the basis for a hypothesis of microbial adaptation and evolution by mutational activation of cryptic genes. Evidence is presented, and a mathematical model is derived, indicating that powerful and biologically important mechanisms exist to prevent the loss of cryptic genes. We propose that cryptic genes persist as a vital element of the genetic repertoire, ready for recall by mutational activation in future generations. Cryptic genes provide a versatile endogenous genetic reservoir that enhances the adaptive potential of a species by a mechanism that is independent of genetic exchange.

Introduction

Recent advances in nucleic acid technology have led to the detection in higher eukaryotes of DNA sequences that appear to be phenotypically silent. Because some of these silent sequences may spread within the genome by forming additional copies of themselves, and because they make no obvious contribution to the fitness of the host organism, they have been termed "selfish DNA" (Dawkins 1976; Doolittle and Sapienza 1980; Orgel and Crick 1980). Another class of noncoding DNA called pseudogenes is defined as a region of DNA that is homologous to a coding sequence but which contains mutations that would prevent its expression (Jacq et al. 1977; Lauer et al. 1980; Nishioka et al. 1980; Li et al. 1981).

The detection of pseudogenes and the concept of selfish DNA have resulted from the application of recombinant DNA technology to studies of higher eukaryotes. Silent genes are also present among prokaryotic and eukaryotic microorganisms, but in most cases their existence was not initially revealed by direct examination of the genome. Instead, these genes were first detected as the result of phenotypic changes that occurred when the silent, or "cryptic," genes were

1. Key words: microorganisms, cryptic genes, adaptation.

Address for correspondence and reprints: Barry G. Hall, Biological Sciences Group, U-44, University of Connecticut, Storrs, Connecticut 06268.

Mol. Biol. Evol. 1(1):109–124.

0737-4038/83/0101-0002$02.00

reactivated. We define "cryptic genes" as phenotypically silent DNA sequences not normally expressed during the life cycle of an individual but capable of activation as a rare event in a few members of a large population by mutation, recombination, insertion elements, or other genetic mechanisms.

In this report we first describe two cases in which the molecular basis for the activation of a cryptic gene is understood. We next review what is known about the activation of other cryptic genes in the context of short-term adaptations and long-term evolutionary development. A unifying hypothesis of the evolutionary significance of cryptic genes in microorganisms is presented.

Activation of the *ilvG* Pseudogene of *Escherichia coli* K-12

The *ilvGEDA* transcription unit of *Escherichia coli* K-12 has recently been shown to contain a naturally occurring frameshift site near the middle of the *ilvG* gene (Lawther et al. 1981, 1982). Mutant derivatives with one-base-pair deletions or two-base-pair insertions that correct the translational alignment of the *ilvG* gene are readily isolated. The wild-type cell produces an amino terminal fragment of the *ilvG* gene product that appears to be rapidly degraded. The mutant *ilvG* gene codes for α-acetohydroxyacid synthase II, an isozyme that differs in some respects from isozymes I and III that are coded by unlinked genes (Bachman and Low 1980) in separate transcription units. Isozymes I and III, but not II, are sensitive to inhibition of catalytic activity by the pathway's end product, valine (Umbarger 1978; DeFelice et al. 1979). For this reason, the wild-type and mutant cells are readily distinguished owing to the valine-resistant growth phenotype of the mutants expressing the valine-resistant *ilvG* coded α-acetohydroxyacid synthase II.

The *ilvG* isozyme II is not required under most growth conditions, since isozymes I and III have adequate biosynthetic capacities. In fact, the cells containing a mutationally activated *ilvG* gene overproduce the *ilvEDA* gene products and wastefully excrete valine into the medium (Rowley 1953; Leavitt and Umbarger 1962). However, when a moderate concentration of valine (1 μg/ml or more) is present in the medium, the growth of the wild-type cell, but not the mutant, is inhibited. Thus, whether the wild-type or the mutant cells exhibit the advantageous phenotype depends on the environmental circumstances. Since transcription of the *ilvGEDA* genes is from *ilvG* to *ilvA* (Berg et al. 1979; Subrahmanyam et al. 1980), the naturally occurring frameshift site in *ilvG* is polar on, and therefore reduces, *ilvEDA* expression. Since the *ilvA* gene product is an end product inhibitable allosteric enzyme which may participate in control of *ilvGEDA* expression (Calhoun and Hatfield 1975), the effects of the naturally occurring frameshift mutation in *ilvG* are amplified.

In a survey of over 300 independent isolates of *E. coli,* three valine-sensitive and 353 valine-resistant strains were observed (Rowley 1953). This observation excludes a trivial explanation, namely, that the *E. coli* K-12 valine-sensitive phenotype is artifactual, arising only from a mutation acquired in the laboratory. The notion that microbial "mutants" are laboratory freaks is firmly ingrained in our thought processes. The possibility that mutant types isolated in the laboratory might accurately resemble strains occurring either commonly or infrequently in nature is not usually given much consideration. And yet, genetic diversity is the hallmark of microbial populations. The distinction between "wild type" (by definition the beginning strain, typically taken directly from nature), and "mutant"

(by definition derived in the laboratory from the wild type) is semantically essential. The discussion at hand is more than semantic, however. In this case the valine-resistant "mutant" strain resembles the more common natural isolate. The phenotypic variability of the *ilvG* gene observed in the laboratory parallels the phenotypic variability observed in strains freshly isolated from nature. It is, in addition, similar to the occurrence in nature of mutants and wild types among many species and for many genetic systems, as discussed in detail below.

In summary, the presence of the naturally occurring frameshift in the wild type *E. coli* K-12 *ilvG* gene has potentially beneficial consequences, including (i) elimination of the *ilvG* coded isozyme II that is not essential under many growth circumstances; (ii) the efficient allosteric control by valine, a pathway end product, since isozymes I and III are inhibited by valine; (iii) expression of *ilvEDA* at less than maximum levels that are, nevertheless, adequate for maximal growth rate; (iv) altered levels of production of the *ilvA* gene product, threonine deaminase, a protein with both allosteric and genetic regulatory properties; and (v) production of pathway end products at levels that are ample for growth but below levels that lead to futile overproduction and excretion. When the *ilvG* pseudogene carried by the wild-type *E. coli* K-12 cells is activated by spontaneous mutations, isozyme II is produced and the cell is phenotypically valine resistant. This selective advantage in the presence of valine is at the expense of economical and balanced amino acid synthesis in the absence of valine as detailed in ii–v above.

Activation of the *bglBSRC* Operon of *Escherichia coli* K-12

In several cases cryptic genes have been identified when bacterial species have acquired new metabolic functions by mutation.

A well-studied example is the phospho-β-glucosidase system of the family Enterobacteriaceae. The β-glucosides are metabolized by a complex system of permeases and phospho-β-glucosidases with different substrate specificities. *Klebsiella* species possess a complete set of permeases and hydrolases and consequently metabolize the aryl β-glucosides arbutin and salicin and the disaccharide cellobiose. Most *Citrobacter* strains utilize cellobiose but do not utilize the aryl β-glucosides, while most *Proteus vulgaris* strains utilize arbutin and salicin but not cellobiose. *Salmonella* and *Escherichia coli* are unable to utilize any β-glucoside sugars (Schaefler and Mintzer 1959; Schaefler and Malamy 1969). Classically these phenotypes would be explained by the loss of genetic information: partial loss in the case of *Proteus* and *Citrobacter* and complete loss in the cases of *Salmonella* and *E. coli*. Both *Salmonella* and *E. coli,* however, mutate easily to β-glucoside positive phenotypes, in the case of *E. coli* at spontaneous frequencies as high as 10^{-5} (Schaefler and Mintzer 1959; Schaefler and Schenkien 1968; Reynolds et al. 1981). Interestingly, the β-glucoside positive mutants do not acquire the ability to utilize the full range of β-glucoside sugars. Mutants of *Salmonella* utilize cellobiose but not the aryl β-glucosides, while mutants of *E. coli* utilize arbutin and salicin but not cellobiose (Schaefler and Mintzer 1959; Schaefler and Schenkien 1968; Schaefler and Malamy 1969; Reynolds et al. 1981).

The β-glucoside system was initially investigated in *E. coli* K-12 by Schaefler and his colleagues (Schaefler 1967; Schaefler and Maas 1967; Prasad and Schaefler 1974) and has recently been investigated by others (Defez and DeFelice 1981; Reynolds et al. 1981; DiNardo et al. 1982). The gene for phosphoglucosidase A (*bglA*) is expressed constitutively in wild-type cells, but the genes for phospho-

glucosidase B (*bglB*), for the transport system (*bglC*), and for the β-glucoside dependent positive regulatory protein (*bglS*) are not expressed. Since the hydrolytic enzymes act only on phosphorylated β-glucosides, only those cells that express the phosphoenolpyruvate dependent β-glucoside transport system specified by the *bglC* gene can utilize β-glucosides. Mutations in the cis-acting *bglR* site are required for the inducible expression of the *bglBSRC* operon. It appears that the *bglR* mutations create a site for the activation of transcription of the operon (Prasad and Schaefler 1974). Spontaneous *bglR* mutations arise at a frequency of about 10^{-5} and are the consequence of integration of insertion sequence IS1 or IS5 into a specific region of the chromosome. Two models may explain the insertional activation of the *bgl* operon (Defez and DeFelice 1981; Reynolds et al. 1981). According to the first model, the operon lacks a functional promoter and the insertion elements contribute to the formation of a promoter. Neither of these elements is thought to contain a recognizable promoter itself; however, this conclusion has recently been questioned (Ciampi et al. 1982). According to the second model, the operon contains an operator site which is disrupted by the insertions. This model implies the existence of a repressor which is refractory to β-glucosides as inducers. The second model is supported by the discovery of a class of mutations which are sufficient to activate the operon even in cells with a wild-type *bglR* gene (Defez and DeFelice 1981). These mutations are located at 27 min on the *E. coli* map, on the opposite side of the chromosome from the *bgl* operon (83 min). Two such mutations were shown to be amber mutations and to be recessive to the wild-type allele. The new gene is designated *bglY*, and the authors suggest that it specifies a repressor which prevents transcription of the *bgl* operon.

The *bgl* operon is thus cryptic, but it can be activated by mutations in a variety of sites. In addition to mutations in *bglR* and *bglY*, it has recently been reported that the operon can also be activated by mutations in *gyrA* and *gyrB*, the genes for DNA gyrase (DiNardo et al. 1982). The authors suggest that expression of the *bgl* operon may be controlled by the degree of local supercoiling of the DNA.

The cryptic β-glucoside system is not limited to laboratory strains of *E. coli*. When 17 "wild" *E. coli* strains, isolated from both humans and a variety of animals from several locales in the United States (Selander and Levin 1980), were tested, all were unable to utilize β-glucosides. All of those strains, many of which had been maintained at −70 since their isolation, yielded spontaneous arbutin-utilizing mutants (Hall, unpublished observations). This observation, together with earlier reports (Schaefler and Schenkien 1968; Schaefler and Malamy 1969), indicates that the cryptic gene was present at the time of, and has been retained since, the evolutionary divergence that led to the modern *E. coli* species.

Naturally Occurring Auxotrophs

The serological and fermentation patterns exhibited by clinical isolates of *Neisseria gonorrhoeae* are adequate to identify the species but inadequate for detailed epidemiological studies. It was observed, however, that the nutritional growth requirements for amino acids, vitamins, purines, and pyrimidines vary widely among clinical isolates. Accordingly an "auxotyping" method based on these differing growth requirements is currently used to identify strains that are indistinguishable by other methods (Carfio and Catlin 1973; Eisenstein et al. 1977;

Knapp et al. 1978; Juni and Heym 1980). It is of interest that "revertants" to prototrophy for most of these growth factors (e.g., isoleucine, thiamine pyrophosphate) are observed in the laboratory (Carfio and Catlin 1973; Catlin 1973; Eisenstein et al. 1977; Knapp et al. 1978; Juni and Heym 1980). The occurrence of these revertants provides evidence that the essential genes were present in a cryptic state, inactivated by mutation. The auxotype of clinical isolates is almost always stable during repeated laboratory passage and during human infections and human-to-human transmission. So, even though a few revertants are commonly observed, the predominant auxotype initially present is only rarely observed to vary in nature.

The genus *Lactobacillus* is well known for its multiple nutritional requirements, including amino acids, purines, pyrimidines, and vitamins. In a study of the genetic basis for these requirements, Morishita et al. (1974) began with a strain of *L. casei* that required 12 amino acids and four vitamins, and they isolated mutants that could grow in the absence of a specific nutrient. They were able to isolate such mutants with respect to seven of the 12 amino acids and three of the four vitamins at frequencies expected for single-step mutations. Similar results were obtained with a strain of *L. acidophilus,* indicating the generality of the phenomenon. The nature of the mutations is unknown, but their existence makes it clear that the original auxotrophy did not result from multiple lesions or from irreversible loss of information via deletions.

Reversion of naturally occurring auxotrophic characters has been seen with other microbial species, including *Salmonella* sp. (Lederberg 1947) (several amino acids) and *Pasteurella pestis* (Englesberg and Ingram 1957) (several amino acids). The tryptophan requirement of *Shigella dysenteriae* was demonstrated by DNA sequencing to be due to a "down" promoter mutation (Miozzari and Yanofsky 1978) and two mutations (detected genetically) in the *trpE* structural gene coding for anthranilate synthase (Manson and Yanofsky 1976). The remainder of the *S. dysenteriae trp* operon was shown to be functionally intact (Manson and Yanofsky 1976).

Acquisition of New Metabolic Capabilities

Metabolic phenotypes play a key role in the taxonomy of microorganisms, and one of the classical properties of *Escherichia coli* is its inability to utilize citric acid as a sole carbon and energy source (Bergey's manual 1974). Although citrate-utilizing *E. coli* have occasionally been isolated from nature (Isiguro et al. 1978), the genes for citrate utilization have invariably been plasmid borne, usually on drug resistance plasmids. A spontaneous citrate-utilizing mutant of *E. coli* K12 has recently been isolated (Hall 1982*a*). Citrate utilization requires mutations in two chromosomal genes, one on each side of the *gal* operon; and the mutant possesses a semiconstitutive citrate transport system. The properties of that transport system are quite distinct from those of at least one plasmid specified citrate transport system (C. Reynolds, personal communication). The parental strain, like other *E. coli,* is unable to transport citrate. Thus, the information required for a citrate transport system has probably been retained in a cryptic form since *E. coli* diverged from its Cit^+ progenitor.

Klebsiella sp. normally possess two *lac* operons, one chromosomal and the other plasmid borne (Reeve and Braithwaite 1974). A mutant strain that had irreversibly lost both of its *lac* operons mutated spontaneously to Lac^+ (Hall

1979). In the mutant strain lactose utilization involved a phosphoenolpyruvate-dependent transport system (Imai and Hall 1981), with the resulting lactose-6-phosphate being metabolized by a phospho-β-galactosidase (Hall 1979). It now appears that the critical mutation to Lac^+ involved decryptification of a gene specifying a lactose specific enzyme II of the PEP-dependent transport system (Hall et al. 1982).

The cases above should not be confused with cases in which new metabolic capabilities arose via point mutations in regulatory genes or via point mutations that altered the catalytic specificities of enzymes (Clarke 1978; Hall 1983).

Retention of Cryptic Genes

The phenomenon of cryptic genes seems to be very general. It seems likely that most microbial species carry genes that are not expressed at a physiologically functional level and which therefore do not make a positive contribution to fitness. It is expected that mutations in such genes will eventually lead to permanent inactivation (Dykhuizen 1978). Depending on the length of time since the gene became cryptic, an increasing portion of the population is expected to lose the ability to decryptify the fully functional gene. Indeed, there is an example of such a scenario: *Shigella dysenteriae,* a close relative of *E. coli,* is typically Lac^- owing to absence of the gene for a lactose permease (*lacY* gene). It does possess a *lacZ* (β-galactosidase) gene that is homologous with the *lacZ* gene of *E. coli,* and a fully functional repressor (*lacI*) gene. In the absence of the permease gene, the β-galactosidase gene plays no functional role in lactose utilization, and mutations in *lacZ* are expected to accumulate. In fact, the *lacZ* gene of *S. dysenteriae* produces levels of β-galactosidase from 0% to 20% of that of *E. coli,* and the enzymes are more thermolabile than *E. coli* β-galactosidase (Luria 1965).

In view of the above we might ask why cryptic genes are retained within populations in functional forms. (The *bglC* gene has apparently been cryptic since the divergence of *E. coli,* yet a majority of the population has clearly retained the functional gene.)

Several possible mechanisms for retention are considered below, including (i) a growth advantage that derives from a superior metabolic regulatory system; (ii) occasional strong environmental selection for population members that express a cryptic gene, leading to survival only of population members that have retained the cryptic gene in a functional form; (iii) an unknown advantage to the cryptic state as demonstrated for auxotrophy; (iv) toxic intermediates that are present in special circumstances, for example, *galE* mutants, unless another gene is made cryptic; and (v) the possible role of the cryptic gene as a metabolic control mechanism that operates at the level of a population of cells rather than within each cell of a population.

Potential Advantages Due to a Cryptic Gene

The specific detailed biochemical and genetic information available for the cryptic *ilvG* gene provides evidence indicating the selective advantages of the cryptic and activated states of *ilvG* depending on environmental conditions. It seems clear that it is advantageous for a population of *Escherichia coli* to retain the *ilvG* pseudogene for activation during conditions of valine-induced growth inhibition. The spontaneous frequency of this mutational activation (10^{-7}–10^{-8}) is high enough to ensure that *E. coli* populations of ordinary size would include

members that express the *ilvG* coded isozyme. The disadvantages of expressing *ilvG* would appear to favor a return to the cryptic state in the absence of valine. The observation that the majority of natural isolates are valine resistant, and thus presumably express the *ilvG* gene, may indicate that under natural conditions the equilibrium favors the active state; or it may well be that the populations tested are not at equilibrium with respect to this trait and that the frequency of the cryptic gene is currently increasing. In either case, the frequency of valine-sensitive strains in nature is high enough ($\approx 1\%$) to indicate that neither the cryptic nor the active state of *ilvG* is simply a laboratory artifact. Cryptification of the active *ilvG* gene in response to an environmental selection pressure is not merely an assumption. Selection for mutants without *ilvG* activity (by counterselection against the valine-resistant phenotype) yielded two types (Smith et al. 1979). One type has high *ilvEDA* expression at levels identical to the parent and therefore had nonpolar *ilvG* mutations. The other type had low *ilvEDA* expression at levels equal to the wild-type K-12 strain containing the pseudogene and can be presumed to have had polar mutations, as is true of the original *ilvG* pseudogene.

The case of the *Lactobacilli* provides perhaps the best example of the retention of nonfunctional genes. The evidence suggests that the genes for many of the biosynthetic pathways for amino acids, purines, and so on contain lesions in the form of *single* point mutations. Under these conditions the remaining genes in the defective pathways perform no useful function and should be lost as additional mutations occur. For one such pathway such mutations have indeed accumulated: strict tryptophan prototrophs could not be isolated from *L. casei*; however, it was possible to isolate mutants that used the precursors anthranilate or indole (Morishita et al. 1974). This suggests that there were multiple or irreversible lesions prior to anthranilate synthesis. How have the other genes escaped inactivation? We propose that periodically in the history of the population there has been strong selection for decryptification of the genes or pathways described. This selection would presumably be similar to the laboratory selection that revealed their existence. During such periods those members of the population that had accumulated inactivating mutations in the now-vital genes would be lost.

If there is occasional strong selection for the expression of normally cryptic genes, we should expect that the genes would remain decryptified even after those selective conditions have ceased to exist. We are thus confronted by an apparent paradox: if there is occasional selection for expression of "cryptic" genes, why are they cryptic at all? To resolve this paradox it is necessary to propose strong selection for cryptification of those genes under circumstances where the gene product is not required.

Formally, then, we propose that under one set of conditions members of the population with a cryptic gene are more fit than those members who express the gene in question, while under some alternative set of conditions those members who express the gene are at a strong selective advantage.

In addition to the intuitive basis above, there is a mathematical basis for the proposal that some type of selective advantage is operating to retain cryptic genes in the population.

Equilibrium Frequencies of Cryptic Genes

We distinguish three classes of genes in a haploid population: (i) cryptic genes, A_1; (ii) functional genes, A_2; and (iii) nonfunctional or irreversibly inactivated

genes, A_3. Let us denote the frequencies of A_1, A_2, and A_3 by x_1, x_2, and x_3, respectively. Assume that the rates of mutation from A_1 to A_2 and from A_2 to A_1 in each generation are v_1 and v_2, respectively. Furthermore, we assume that the rates of irreversible mutation from A_2 to A_3 and from A_1 to A_3 are the same and μ per generation. Let 1, $1 - s$, and $1 - t$ be the relative fitnesses of A_1, A_2, and A_3, respectively. Thus, we assume constant fitness. As already noted, the fitness may vary from time to time owing to environmental change. Under that condition, the fitness in the present model can be interpreted as the mean values of the fitnesses through time.

After the processes of mutation and selection, the changes in the gene frequencies are given by

$$\begin{aligned}\Delta x_1 &= x_1(sx_2 + tx_3)/W - (\mu + v_1)x_1 + v_2x_2,\\ \Delta x_2 &= x_2[-s(1-x_2) + tx_3]/W + v_1x_1 - (\mu + v_2)x_2,\\ \Delta x_3 &= x_3[sx_2 - t(1-x_3)]/W + \mu x_1 + \mu x_2,\end{aligned} \tag{1}$$

where $W = 1 - sx_2 - tx_3$.

Before going into detail, we should note that if the frequency of nonfunctional genes in a population becomes unity, no further evolutionary changes occur unless there is migration or other forces which introduce functional or cryptic alleles into that population. Thus, it is important to study the existence of nontrivial equilibria and stability of these points.

In an equilibrium population, it is sufficient to consider the following two equations:

$$\hat{x}_1(s\hat{x}_2 + t\hat{x}_3)/\hat{W} - (\mu + v_1)\hat{x}_1 + v_2\hat{x}_2 = 0\,, \tag{2a}$$

and

$$\hat{x}_3[s\hat{x}_2 - t(1 - \hat{x}_3)]/\hat{W} + \mu(\hat{x}_1 + \hat{x}_2) = 0\,, \tag{2b}$$

where $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, and $\hat{W}$ are the equilibrium values of x_1, x_2, x_3, and W, respectively.

When $s = t = 0$, it is clear that $\hat{x}_1 = \hat{x}_2 = 0$, and $\hat{x}_3 = 1$ from equations (2a) and (2b). Thus, without selection, nonfunctional genes will eventually fix in the population.

To solve the equations (2a) and (2b), it is convenient to write (2b) as

$$(1 - \hat{x}_3)[\mu - \hat{x}_3(t - \hat{a})/\hat{W}] = 0\,, \tag{3}$$

where $\hat{a} = s\hat{x}_2/(\hat{x}_1 + \hat{x}_2)$. Thus, two sets of equilibrium gene frequencies exist. One is the trivial equilibrium point $\hat{x}_3 = 1$. A nontrivial equilibrium gene frequency can be obtained by solving

$$\mu - (t - \hat{a})\hat{x}_3/\hat{W} = 0\,, \tag{4}$$

where $\hat{W} = 1 - s\hat{x}_2 - t\hat{x}_3$. Putting (4) into (2a), we obtain

$$s\hat{p}\hat{q}/\hat{W} - v_1\hat{p} + v_2\hat{q} = 0\,, \tag{5}$$

where $\hat{p} = \hat{x}_1/(\hat{x}_1 + \hat{x}_2)$ and $\hat{q} = \hat{x}_2/(\hat{x}_1 + \hat{x}_2)$. Furthermore, by using equation (4) and definitions of $\hat{a}$, $\hat{p}$, and $\hat{q}$, we can reduce $\hat{W}$ to

$$\hat{W} = (1 - s\hat{q})/(1 + \mu)\,. \tag{6}$$

From (5) and (6),

$$A\hat{q}^2 - B\hat{q} + v_1 = 0 , \tag{7}$$

where $A = s(1 + \mu + v_1 + v_2)$ and $B = s(1 + \mu + v_1) + v_1 + v_2$. Thus, with the condition that $0 < \hat{q} < 1$,

$$\hat{q} = \begin{cases} v_1/(v_1 + v_2) & \text{if } s = 0, \quad (8a) \\ (B - \sqrt{B^2 - 4v_1A})/(2A) & \text{if } s \neq 0. \quad (8b) \end{cases}$$

Once we obtain $\hat{q}$, from (4),

$$\hat{x}_3 = \mu(1 - s\hat{q})/[(1 + \mu)(t - s\hat{q})] . \tag{9}$$

Then, with the definitions of $\hat{p}$ and $\hat{q}$,

$$\hat{x}_1 = (1 - \hat{q})(1 - \hat{x}_3) , \tag{10}$$

and

$$\hat{x}_2 = \hat{q}(1 - \hat{x}_3) . \tag{11}$$

From (8a) to (11), we can derive nontrivial equilibrium gene frequencies depending on the values of s. Note that there is only one equilibrium point. When $s = 0$ and $t > \mu/(1 + \mu)$,

$$\begin{aligned} \hat{x}_1 &= v_2/(v_1 + v_2)\{1 - \mu/[(1 + \mu)t]\}, \\ \hat{x}_2 &= v_1/(v_1 + v_2)\{1 - \mu/[(1 + \mu)t]\}, \\ \hat{x}_3 &= \mu/[(1 + \mu)t]. \end{aligned} \tag{12}$$

Thus, when there is no selective difference between the functional and cryptic genes but these genes have higher fitness than nonfunctional genes, and when $t > \mu/(1 + \mu)$, then nontrivial equilibrium frequencies of cryptic genes depend mainly on the forward and backward mutation rates between the functional and cryptic genes. Under this condition, the frequency of cryptic genes can be close to unity if the mutation rate from A_2 to A_1 is much higher than that from A_1 to A_2.

It may be useful to study formulas (8a)–(11) under special circumstances for $s \neq 0$. When s and $t >> \mu$, v_1, and v_2, then $\hat{q} \simeq v_1/s$, and, therefore, $\hat{x}_1 \simeq (1 - v_1/s)(1 - \mu/t)$, $\hat{x}_2 \simeq (v_1/s)(1 - \mu/t)$, and $\hat{x}_3 \simeq \mu/t$. Thus when cryptic genes have higher fitness than functional and nonfunctional genes, the frequency of the cryptic genes is given by $(1 - v_1/s)(1 - \mu/t)$, and this can be close to unity, as expected. When $s \gg t$, μ, v_1, and v_2, then $\hat{x}_1 \simeq (1 - v_1/s)[1 - \mu/(t - v_1)]$, $\hat{x}_2 \simeq (v_1/s)[1 - \mu/(t - v_1)]$, and $\hat{x}_3 \simeq \mu/(t - v_1)$ as long as $t > v_1$. Thus, as the value of t gets smaller, the frequency of cryptic genes becomes smaller. Let us now consider the case of $t = 0$, that is, the case where there is no selective difference between cryptic and nonfunctional genes. In this case, equation (4) reduces to $\mu + \hat{a}\hat{x}_3/\hat{W} = 0$, or $\mu + s\hat{x}_2\hat{x}_3/[(\hat{x}_1 + \hat{x}_2)\hat{W}] = 0$. Thus, when $s > 0$, then there is no nontrivial equilibrium point. When $s < 0$, however, a nontrivial equilibrium frequency of cryptic genes exists. For example, by setting $s' = -s$, $\hat{x}_1 \simeq (v_2/s')[1 - \mu(1 + s')/s']$, $\hat{x}_2 \simeq (1 - v_2/s')[1 - \mu(1 + s')/s']$, and $\hat{x}_3 \simeq \mu(1 + s')/s'$ as long as $s' > \mu/(1 - \mu)$.

Some numerical examples of the equilibrium gene frequencies are given in table 1, where $\mu = v_1 = v_2 = 10^{-5}$ were used. These numerical examples show that the frequency of cryptic genes becomes high when the values of s and t are large, as expected. We should, however, point out that even when selection coefficients s and t differ by only an order of magnitude from the mutation rates, the frequency of cryptic genes is still high. For example, when $s = t = 10^{-4}$ and $\mu = v_1 = v_2 = 10^{-5}$, then $\hat{x}_1 \simeq 0.81$ and $\hat{x}_3 = 0.11$ (see table 1). When $s = 0$ and $t > 0$, the frequency of cryptic genes is given approximately by $v_2/(v_1 + v_2)$ (see eq. [12]). Thus, if $v_1 = v_2$, that frequency is 0.5 and still substantial. Although the frequency of cryptic genes is much lower for $s < 0$ (i.e., cryptic alleles are less fit than functional alleles), cryptic alleles do persist in the population. For example, the values of x_1 are given by 0.123, 0.080, and 0.010 for $s = -0.00005$, -0.0001, and -0.001, respectively (table 1).

Stability of the equilibrium values in table 1 can be studied numerically using equation (1). The analyses show that these equilibrium points are locally stable. Thus, when the gene frequencies deviate from an equilibrium point they will return to that point. For example, consider the case of $s = 0.01$ and $t = 0.001$ with $\mu = v_1 = v_2 = 10^{-5}$. Suppose that $x_1 = 0.01$, $x_2 = 0.9$, and $x_3 = 0.09$ in the initial population, then $(x_1, x_2, x_3) = (0.1494, 0.0834, 0.7671)$, $(0.9316, 0.0008, 0.0674)$, and $(0.9885, 0.0010, 0.0105)$ after 500, 5,000, and 10,000 generations. At generation 12,065, $x_1 = 0.9889$, $x_2 = 0.0010$, and $x_3 = 0.0101$, which is very close to the equilibrium values given in table 1.

These deterministic analyses show that the frequency of cryptic genes can be close to unity even when the cryptic genes have only slightly higher fitnesses than the functional genes. The analyses also show that all three classes of genes can be maintained even if the functional genes have higher fitness than the cryptic and nonfunctional genes.

Table 1
Numerical Examples of Equilibrium Gene Frequencies

s	t	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$
.01	.01	.9980	.0010	.0010
	.001	.9889	.0010	.0101
	.0001	.8880	.0009	.1111
	.0005	.7493	.0008	.2499
	.00003	.4997	.0006	.4997
.001	.01	.9890	.0100	.0010
	.001	.9801	.0099	.0100
	.001	.8802	.0088	.1110
.0001	.01	.9110	.0880	.001
	.001	.9007	.0892	.0101
	.0001	.8099	.0802	.1099
	.00005	.6879	.0681	.2440
	.0003	.4764	.0472	.4764
−.00005	.0	.1230	.6385	.2385
−.0001	.0	.082	.8099	.1099
−.001	,0	.0098	.9801	.0101

NOTE.—In the computation $\mu = v_1 = v_2 = 10^{-5}$ were used.

The present mathematical analyses and the available data strongly suggest that some types of selective advantage of the cryptic gene must be operating over functional genes. This theoretical treatment is supported by a number of concrete observations.

There are examples of selection against organisms that express unneeded functions, although most are anecdotal. Morishita et al. (1974) point out that when the mutant strains of *Lactobacillus casei* that had become independent of various nutrients were grown for many generations on a complex medium, they often reverted to the original phenotype. Likewise, Baumann and Baumann (1981) have pointed out that many *Vibrio* species acquire nutritional requirements when stored on complex media. Beckwith (personal communication) has found that when Lac constitutive strains of *E. coli* are stored for long periods on rich medium they accumulate point mutations in the *lacZ* gene. Indeed, it is a common word-of-mouth observation that bacteria which grew quite well on minimal media when originally isolated from nature often acquire auxotrophies when stored or repeatedly transferred on a rich medium.

The Advantage of Auxotrophy

In 1944 Lwoff predicted that in the presence of the end product an auxotroph would have a selective advantage over the prototroph (Lwoff 1944). Later, Zamenhoff and Eichorn (1967) studied this question by competing prototrophs with auxotrophic mutants of *Bacillus subtilis* in glucose-limited chemostats. They found that in the presence of the required amino acid, histidine and tryptophan auxotrophs had a strong selective advantage over prototrophs. They attributed the selective advantage of auxotrophy to energy and resource conservation, but in 1978 Dykhuizen (1978) cast serious doubt on this explanation for auxotrophic advantage. Placing tryptophan auxotrophs of *Escherichia coli* in competition with isogenic prototrophs in glucose-limited chemostats, he also found that auxotrophs had a strong selective advantage. He calculated the percentage of the energy budget devoted to tryptophan biosynthesis in repressed feedback-inhibited cells as 0.01%. He thus predicted, first, that if energy saving is the basis of auxotrophic advantage, the selective advantage should not exceed 0.01%. Second, he predicted that a nonsense mutation in *trpE*, which permits synthesis of neither tryptophan nor any of the *trp* proteins, should have an advantage over a *trpE* missense mutation which permits synthesis of the proteins but not tryptophan. Third, he predicted that the selective advantage of the auxotroph should be greatest when supplied with tryptophan, less when supplied with indole, and least when supplied with anthranilate. When tested by appropriate competition experiments in chemostats, none of the predictions was borne out: the selection for the auxotrophs was three orders of magnitude greater than could be explained on the basis of the energy budget; the selective advantage of the polar mutations was indistinguishable from that of the missense mutations; and there was no difference in the selection for the missense mutant on indole or anthranilate. In short, all of the tests failed to support the energy conservation hypothesis.

Although it does not appear to be based on energy conservation, there is a clear and consistent selective advantage to auxotrophy under conditions where the nutrient is in the environment. How, then, might auxotrophy be advantageous? We can imagine that some intermediates in anabolic pathways might be slightly toxic or might slightly perturb the regulation of some other pathway. The presence

of such an intermediate when the pathway end product is not required would slightly reduce growth rates below those that would be achieved in a mutant that no longer expressed the pathway.

An example from a laboratory situation suggests that in some circumstances it could be advantageous to cryptify catabolic genes. *Escherichia coli* mutants that are defective in the *galE* gene, the gene for UDP-galactose epimerase, are killed by galactose owing to the accumulation of highly toxic UDP-galactose (Ippen et al. 1971). In *galE*$^-$ strains, lactose negative and melibiose negative mutations are strongly selected in the presence of lactose or melibiose because metabolism of these substrates results in the generation of internal galactose. Although this could be viewed as simply an artifactual laboratory situation, it must be realized that the wide variety of metabolic capabilities that we now regard as "wild type" and as defining the various bacterial species might well be viewed as genetic deficiencies if we were aware of the phenotypes of the ancestral organisms from which they derived.

Another possibility is that an unneeded protein might act directly to reduce fitness by interacting inappropriately with other cellular components. A transport protein that competed for a limited number of sites in the membrane could easily reduce fitness when it was no longer required. Dykhuizen and Davies (1980) have shown that lactose use interferes with the concurrent use of maltose and have provided evidence supporting the hypothesis that this interference is due to competition for a limited number of membrane sites by the lactose and maltose permeases. As another example, the lactose permease has been implicated in "lactose killing" under unusual environmental circumstances (Dykhuizen and Hartl 1978).

If expression of a gene could reduce fitness, why wouldn't the organism simply evolve a regulatory system to prevent expression of that gene when it was not required? The results of various directed evolution studies (reviewed in Clark [1978]; Hall [1982*b*, 1983]) suggest that microorganisms do not evolve "optimal" solutions to problems but simply take the first solution that solves the problem. For many systems it is probably far easier to cryptify a gene than it is to evolve a regulatory system.

Cryptic Genes as a Metabolic Control System

The activation of cryptic genes by mutation is a regulatory event that permits the expression of an otherwise silent gene. Other mechanisms that regulate gene and enzyme activity, such as induction, repression, and allosteric control, are most effective for those circumstances requiring frequent or continuous modulation during the lifetime of most individual members of a species. This physiological regulation generates a homogeneous adaptation to an environmental change; that is, the genetic constitution of the population remains unchanged. Cryptic gene activation operates at a higher hierarchical level, affecting only a very few members of a species. Environmental changes may occur in a repetitive fashion over long evolutionary time spans. Indeed, such changes typify the evolutionary history of many extant species. We view cryptification and decryptification of genes as adaptations to these cyclic environmental changes. These adaptations occur more frequently than, and are superimposed on, long-term evolutionary changes that may occur during periods of relative stasis (Cronin et al. 1981). One consequence of an environmental change that selects for expression of a cryptic gene would be a sharp reduction in the genetic variability of the population, similar

to that which occurs during periodic selection in chemostats. Since microbial populations probably enjoy little genetic exchange in natural populations (Selander and Levin 1980), such cyclic changeovers are expected to result in populations with a very limited number of distinctive genotypes. That expectation appears to be fulfilled for natural *Escherichia coli* populations. In a survey of 109 clones from geographically dispersed natural populations, based on electrophoretic mobilities of 20 different enzymes, Selander and Levin (1980) detected three pairs of identical clones isolated from hosts that were separated by wide geographic distances. Since the mean genetic diversity of that population was very high (0.47), they concluded that the natural population of *E. coli* consists of a very limited number of clones. The activation of cryptic genes is, therefore, an event with both regulatory and evolutionary implications. The existence of various mechanisms for activating cryptic genes emphasizes the widespread nature of the phenomenon.

We view our proposed role for cryptic genes as being applicable primarily to microorganisms. First, null alleles are rare in populations of higher eukaryotes. Second, because the somatic tissue provides a buffer between gene transmission and environment, higher eukaryotes are less subject to the boom-or-bust of variation in substrates that affects microorganisms. Eukaryotic microorganisms such as yeasts, however, should be subject to the same selective pressures as those encountered by bacteria, and we fully expect that those species also contain cryptic genes. Consistent with this expectation, Carlson et al. (1981) have shown that *Saccharomyces cerevisiae* strains often possess silent genes for sucrose metabolism and that these alleles can be mutationally activated.

The ubiquity of cryptic gene activation in microorganisms raises questions about the possibility that genes in higher forms may similarly be activated in a small fraction of the individuals of a population. It is noteworthy in this regard that the activation of vertebrate genes as a result of environmental carcinogenesis or by genetic insertion of tumor viruses has recently been documented (Goldfarb et al. 1982; Reddy et al. 1982; Tabin et al. 1982). There is insufficient evidence available at this time, however, to propose that the mechanisms that lead to vertebrate gene activation serve the same positive evolutionary roles as the type we propose for microorganisms. However, there is no evidence to the contrary, and the possibility should not be excluded in the absence of further experimental evidence.

Acknowledgments

We are indebted to L. Sarokin for first raising the question about retention of cryptic genes in bacterial populations, to J. Antonovics for very helpful discussions, and to J. M. Calvo for his particularly insightful comments on an earlier version of this paper. We are also grateful to Wen-Hsiung Li, Bruce Levin, and Fumio Tajima for their comments on the theoretical aspects of this work. Fumio Tajima's contribution on the derivation of exact solutions from equation (1) was particularly important.

Barry G. Hall is supported by Research Career Development Award 1 KO4 AI00366 from the National Institute of Allergy and Infectious Diseases. Research in his laboratory was supported by grants from the National Institutes of Health and the National Science Foundation. Shozo Yokoyama is supported by grants GM28672 and MH31302 from the U.S. Public Health Service with Washington University. David H. Calhoun is the recipient of an Irma T. Hirschl Career Scientist

Award. Research in his laboratory is supported by grant GM23182 from the National Institutes of Health.

LITERATURE CITED

BACHMANN, B. J., and K. B. LOW. 1980. Linkage map of *Escherichia coli* K-12, edition 6. Microbiol. Rev. **44:**1–56.

BAUMANN, P., and L. BAUMANN. 1981. The marine gram-negative eubacteria: genera *Photobacterium, Beneckea, Alteromonas, Pseudomonas,* and *Alcaligenes*. Pp. 1302–1331 *in* M. P. STARR, H. STOLP, H. G. TRUPER, A. BALLOWS, and H. G. SCHLEGEL, eds. The prokaryotes. Springer, Heidelberg.

BERG, C. M., K. J. SHAW, J. VENDER, and M. BORUCKA-MANKIEWICZ. 1979. Physiological characterization of polar Tn5-induced isoleucine-valine auxotrophs in *Escherichia coli* K12: evidence for an internal promoter in the *ilvOGEDA* operon. Genetics **93:**309–319.

Bergey's manual of determinative bacteriology. 8th ed. 1974. Edited by N. E. BUCHANAN and R. E. GIBBONS. Williams & Wilkins, Baltimore.

CALHOUN, D. H., and G. W. HATFIELD. 1975. Autoregulation of gene expression. Annu. Rev. Microbiol. **29:**275–299.

CARFIO, K., and B. W. CATLIN. 1973. *Neisseria gonorrhoeae* auxotyping: differentiation of clinical isolates based on growth responses on chemically defined media. Appl. Microbiol. **26:**223–230.

CARLSON, M., B. C. OSMOND, and D. BOTSTEIN. 1981. Genetic evidence for a silent *SUC* gene in yeast. Genetics **94:**41–54.

CATLIN, B. W. 1973. Nutritional profiles of *Neiseria gonorrhoeae, Neisseria meningitidis,* and *Neisseria lactamica* in chemically defined media and use of growth requirements for gonococcal typing. J. Infect. Dis. **128:**178–194.

CIAMPI, M. S., M. B. SCHMID, and J. R. ROTH. 1982. Transposon Tn10 provides a promoter for transcription of adjacent sequences. Proc. Natl. Acad. Sci. USA **79:**5016–5020.

CLARKE, P. H. 1978. Experiments in microbial evolution. Pp. 137–218 *in* L. N. ORNSTON and J. R. SOKATCH, eds. The bacteria. Vol. **6.** Academic Press, New York.

CRONIN, J. E., N. T. BOAZ, C. B. STRINGER, and Y. RAK. 1981. Tempo and mode in hominid evolution. Nature **292:**113–122.

DAWKINS, R. 1976. The selfish gene. Oxford University Press, Oxford.

DEFELICE, M., M. LEVINTHAL, M. IACCARINO, and J. GUARDIOLA. 1979. Growth inhibition as a consequence of antagonism between related amino acids: effect of valine in *Escherichia coli* K12. Microbiol. Rev. **43:**42–58.

DEFEZ, R., and M. DEFELICE. 1981. Cryptic operon for β-glucoside metabolism in *Escherichia coli* K12: genetic evidence for a regulatory protein. Genetics **97:**11–25.

DINARDO, S., K. A. VOELKEL, R. STERNGLANZ, A. E. REYNOLDS, and A. WRIGHT. 1982. *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell **31:**43–51.

DOOLITTLE, W. F., and C. SAPIENZA. 1980. Selfish genes, the phenotype paradigm, and genome evolution. Nature **284:**601–603.

DYKHUIZEN, D. 1978. Selection for tryptophan auxotrophy of *Escherichia coli* in glucose limited chemostats as a test of the energy conservation hypothesis. Evolution **32:**125–150.

DYKHUIZEN, D., and M. DAVIES. 1980. An experimental model: bacterial specialists and generalists competing in chemostats. Ecology **135:**1213–1227.

DYKHUIZEN, D., and D. L. HARTL. 1978. Transport by the lactose permease of *Escherichia coli* as the basis of lactose killing. J. Bacteriol. **135:**876–882.

EISENSTEIN, B. I., T. S. LEE, and P. F. SPARLING. 1977. Penicillin sensitivity and serum resistance are independent attributes of strains of *Neisseria gonorrhoeae* causing disseminated gonococcal infection. Infect. Immunol. **15:**834–841.

ENGLESBERG, E., and L. INGRAM. 1957. Meitrophic mutants of *Pasteurella pestis* and their use in the elucidation of nutritional requirements. Proc. Natl. Acad. Sci. USA **43:**369–372.

GOLDFARB, M., K. SHIMIZU, M. PERUCHO, and M. WIGLER. 1982. Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. Nature **296:**404–409.

HALL, B. G. 1979. Lactose metabolism involving phospho-β-galactosidase in *Klebsiella*. J. Bacteriol. **138:**691–698.

———. 1982*a*. Chromosomal mutation for citrate utilization in *Escherichia coli* K12. J. Bacteriol. **151:**269–273.

———. 1982*b* Evolution on a petri dish: the evolved β-galactosidase system as a model for studying acquisitive evolution in the laboratory. Evol. Biol. **15:**85–150.

———. 1983. Evolution of new metabolic functions in laboratory organisms. Pp. 234–257 *in* M. NEI and R. K. KOEHN, eds. Evolution of genes and proteins. Sinauer, Sunderland, Mass.

HALL, B. G., K. IMAI, and C. ROMANO. 1982. Genetics of the Lac-PTS system in *Klebsiella*. Genet. Res. **39:**287–302.

IMAI, K., and B. G. HALL. 1981. Properties of the lactose transport system in *Klebsiella* sp. strain CT-1. J. Bacteriol. **145:**1459–1462.

IPPEN, K., J. A. SHAPIRO, and J. R. BECKWITH. 1971. Transposition of the *lac* region to the *gal* region of the *Escherichia coli* chromosome: isolation of λ*lac* transducing bacteriophages. J. Bacteriol. **108:**5–9.

ISHIGURO, N., C. OKA, and G. SATO. 1978. Isolation of citrate positive variants of *Escherichia coli* from domestic pigeons, pigs, cattle, and horses. Appl. Environ. Microbiol. **36:**217–222.

JACQ, C., J. R. MILLER, and G. G. BROWNLEE. 1977. A pseudogene structure in 5S DNA of *Xenopus laevis*. Cell **12:**109–120.

JUNI, E., and G. A. HEYM. 1980. Studies of some naturally occurring auxotrophs of *Neisseria gonorrhoeae*. J. Gen. Microbiol. **121:**85–92.

KNAPP, J. S., C. THORNDYKE, G. A. SCHOOLNI, P. J. WIESNER, and K. K. HOLMES. 1978. Phenotypic and epidemiologic correlation of auxotype in *Neisseria gonorrhoeae*. J. Infect. Dis. **138:**160–165.

LAUER, J., C. K. J. SHEN, and T. MANIATIS. 1980. The chromosomal arrangements of human α-like globin genes: structure homology and α globin gene deletions. Cell **20:**119–130.

LAWTHER, R. P., D. H. CALHOUN, C. W. ADAMS, C. A. HAUSER, J. GRAY, and W. HATFIELD. 1981. Molecular basis of valine resistance in *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA **78:**922–925.

LAWTHER, R. P., D. H. CALHOUN, J. GRAY, C. W. ADAMS, C. M. HAUSER, and G. W. HATFIELD. 1982. DNA fine structure analysis of *ilvG* (IlvG$^+$) mutation of *Escherichia coli* K12. J. Bacteriol. **149:**294–298.

LEAVITT, R. I., and H. E. UMBARGER. 1962. Isoleucine and valine metabolism in *Escherichia coli*. XI. Valine inhibition of the growth of *Escherichia coli* strain K-12. J. Bacteriol. **83:**624–630.

LEDERBERG, J. 1947. The nutrition of *Salmonella*. Arch. Biochem. **13:**287–290.

LI, W. H., T. GOJOBORI, and M. NEI. 1981. Pseudogenes as a paradigm of neutral evolution. Nature **292:**237–239.

LURIA, S. E. 1965. On the evolution of the lactose utilization gene system in enteric bacteria. Pp. 357–358 *in* V. BRYSON and H. J. VOGEL, eds. Evolution of genes and proteins. Academic Press, New York.

LWOFF, A. 1944. L'Evolution physiologique: étude des pertes de fonctions chez les microorganismes. Hermann, Paris.

MANSON, M. D., and C. YANOFSKY. 1976. Naturally occurring sites within the *Shigella*

dysenteriae tryptophan operon severely limit tryptophan biosynthesis. J. Bacteriol. **126:**668–678.

MIOZZARI, G., and C. YANOFSKY. 1978. Naturally occurring promoter down mutation: nucleotide sequence of the trp promoter/operator/leader region of *Shigella dysenteriae* 16. Proc. Natl. Acad. Sci. USA **75:**5580–5584.

MORISHITA, T., T. FUKADA, M. SHIROTA, and T. YURA. 1974. Genetic basis of nutritional requirements in *Lactobacillus casei*. J. Bacteriol. **120:**1078–1084.

NISHIOKA, Y., A. LEDER, and P. LEDER. 1980. Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA **77:**2806–2809.

ORGEL, L. E., and F. H. C. CRICK. 1980. Selfish DNA—the ultimate parasite. Nature **284:**601–603.

PRASAD, I., and S. SCHAEFLER. 1974. Regulation of the β-glucoside system in *Escherichia coli* K12. J. Bacteriol. **120:**638–650.

REDDY, E. P., R. K. REYNOLDS, E. SANTAS, and M. BARBACID. 1982. A point mutation is responsible for the acquisition of transforming properties by T24 human bladder carcinoma. Nature **300:**149–152.

REEVE, E. C. R., and J. A. BRAITHWAITE. 1974. The lactose system of *Klebsiella aerogenes* V9A. 4. A comparison of the Lac operons of *Klebsiella* and *Escherichia coli* K12. Genet. Res. **24:**323–331.

REYNOLDS, A., Y. FELTON, and A. WRIGHT. 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K-12. Nature **293:**625–629.

ROWLEY, D. J. 1953. Interrelationships between amino acids in the growth of coliform organisms. J. Gen. Microbiol. **9:**37–43.

SCHAEFLER, S. 1967. Inducible system for the utilization of β-glucosides in *Escherichia coli*. I. Active transport and utilization of β-glucosides. J. Bacteriol. **93:**254–263.

SCHAEFLER, S., and W. K. MAAS. 1967. Inducible system for the utilization of β-glucosides in *Escherichia coli*. II. Description of mutant types and genetic analysis. J. Bacteriol. **93:**264–272.

SCHAEFLER, S., and A. J. MALAMY. 1969. Taxonomic investigations on expressed and cryptic phospho-β-glucosidases in *Enterobacteriaceae*. J. Bacteriol. **99:**422–433.

SCHAEFLER, S., and L. MINTZER. 1959. Acquisition of lactose fermenting properties by *Salmonellae*. I. Interrelationship between the fermentation of cellobiose and lactose. J. Bacteriol. **78:**159–163.

SCHAEFLER, S., and I. SCHENKEIN. 1968. β-glucoside permeases and phospho-β-glucosidases in *Aerobacter aerogenes:* relationship with cryptic phospho-β-glucosidase in *Enterobacteriaceae*. Proc. Natl. Acad. Sci. USA **59:**285–292.

SELANDER, R. K., and B. R. LEVIN. 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210:**545–547.

SMITH, J. A., F. S. SMITH, and H. E. UMBARGER. 1979. Mutations affecting the formation of acetohydroxy acid synthase II in *Escherichia coli* K12. Mol. Gen. Genet. **169:**299–314.

SUBRAHMANYAM, C. S., J. D. NOTI, and H. E. UMBARGER. 1980. Regulation of *ilvEDA* expression occurs upstream of *ilvG* in *Escherichia coli:* additional evidence for an *ilvGEDA* operon. J. Bacteriol. **144:**279–290.

TABIN, C. J., S. M. BRADLEY, C. I. BARGMANN, R. A. WEINBERG, A. G. PAPAGEORGE, E. M. SCOLNICK, R. DAHR, D. R. LOWY, and E. H. CHANG. 1982. Mechanism of activation of a human oncogene. Nature **300:**143–149.

UMBARGER, H. E. 1978. Amino acid biosynthesis and its regulation. Annu. Rev. Biochem. **47:**533–606.

ZAMENHOFF, S., and H. H. EICHORN. 1967. Studies of microbial evolution through loss of biosynthetic functions: establishment of defective mutants. Nature **216:**455–458.

MASATOSHI NEI, reviewing editor

Received April 18, 1983; revision received July 12, 1983.