Proc. Natl. Acad. Sci. USA
Vol. 78, No. 10, pp. 6344–6348, October 1981
Genetics

Potential for selection among nearly neutral allozymes of 6-phosphogluconate dehydrogenase in *Escherichia coli*

(natural selection/selective neutrality/allozyme/*gnd* gene/chemostat)

DANIEL L. HARTL AND DANIEL E. DYKHUIZEN

Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110

Communicated by Motoo Kimura, June 22, 1981

ABSTRACT **Six *gnd* alleles coding for naturally occurring allozymes of 6-phosphogluconate dehydrogenase [6-phospho-D-gluconate:NAD(P)$^+$ 2-oxidoreductase, EC 1.1.1.43] have been transferred by transduction into the genetic background of *Escherichia coli* K-12 and examined for their selective effects in chemostats in which gluconate was limiting. Four of the alleles are evidently neutral or nearly neutral, inasmuch as their selective effects, if any, fall below the limit of resolution of the procedure—0.5%/hr or about 1% per generation. One allele is detrimental in limiting gluconate but not in limiting glucose or fructose. Another allele has a detrimental, density-dependent, epistatic interaction with *tonA*. We suggest that all six alleles are neutral or nearly neutral in natural populations but that they are not functionally equivalent; their functional differences are potentially important because they can become expressed as differences in fitness under the appropriate conditions of environment or genetic background. Under these conditions, otherwise neutral alleles can become subject to selection.**

Natural populations of *Escherichia coli* contain at least 16 electrophoretically distinct allozymes of the enzyme 6-phosphogluconate dehydrogenase [6PGD; 6-phospho-D-gluconate: NAD(P)$^+$ 2-oxidoreductase, EC 1.1.1.43], which are evidently coded by alternative alleles at the *gnd* locus at 44 min on the standard genetic map (1, 2). The 6PGD polymorphism is typical of many enzymes in *E. coli* and other organisms, but the evolutionary significance of such polymorphisms and the mechanisms of their maintenance in populations are unknown. Two extreme views have become known as the "neutralist hypothesis" and the "selectionist hypothesis." According to the neutralist view, alleles coding for allozymes have so little effect on fitness that their changes in frequency are due primarily to random genetic drift (3, 4). The selectionist view is that the alleles do have effects on fitness and that their frequencies are maintained by a balance of selective forces. There are, of course, many shades of opinion between these extremes.

We have used bacteriophage P1-mediated transduction to transfer allozyme-associated alleles into the genetic background of strain *E. coli* K-12 in order to determine whether these alleles have detectable effects on fitness. Our measure of fitness is the ability of a strain to compete for limiting nutrient in a chemostat. In theory, if $R(t)$ and $S(t)$ represent the relative proportions of two strains competing in a chemostat at a time t hours after inoculation, then $\ln[R(t)/S(t)] = C + st$, in which C is a constant determined by the initial proportions and s is a measure of the relative ability of the strains to compete for the limiting nutrient (5). When $s = 0$, the strains are equal in competitive ability.

Our measure of fitness is the number s. This will vary according to the particular nutrient that limits growth, but maximal amounts of selection are expected when the limiting nutrient is the substrate of the enzyme in question. Because 6PGD generates reducing equivalents in the cell by catalyzing an NADP$^+$-dependent oxidation of 6-phosphogluconate en route to the pentose-phosphate shunt, maximal selection for *gnd* alleles is expected when the limiting nutrient is gluconate, because this compound is phosphorylated by the cell and so becomes the substrate of 6PGD. Any physiologically or evolutionarily relevant differences between 6PGD allozymes would therefore be expected to be expressed or perhaps even amplified under competition for limiting gluconate.

Six *gnd* alleles coding for different allozymes have been studied. Four of these alleles have no detectable effects on fitness under gluconate limitation. The other two alleles are associated with selection of an unexpectedly complex variety.

MATERIALS AND METHODS

Strain Construction. Strains carrying various allozymes were provided by R. D. Milkman and B. R. Levin. Table 1 outlines the protocol used to transfer the *gnd* allele from each strain into a K-12 genetic background. Transduction (6) was mediated by bacteriophage P1 (*cml clr100*), and the 6PGD allozyme was verified electrophoretically after each step. The first two transductions involve the transfer of the *gnd*$^+$–*his*$^+$ region of an original strain into the genetic background of DD715, selection being for *his*$^+$ (44 min) and subsequent scoring for *gnd*. Selection in the third transduction is for growth on gluconate because *edd*$^-$ *gnd*$^-$ cells are unable to grow on gluconate. (*edd* is at 41 min and codes for phosphogluconate dehydratase, EC 4.2.1.12.) In the final transduction, the deletion covering the loci *zwf–edd–eda* is replaced with its normal counterpart from DD657 by selection on glucuronate for *eda*$^+$. (*eda* codes for 6-phospho-2-keto-3-deoxygluconate aldolase, EC 4.1.2.14; *zwf* codes for glucose-6-phosphate dehydrogenase, EC 1.1.1.49.) Each resulting strain thus carries a naturally occurring *gnd* allele in addition to *rpsL*, which confers streptomycin resistance. In the final step, a spontaneous mutation at the *tonA* locus is obtained by selection for resistance to bacteriophage T5. *tonA*$^-$ has proven to be selectively neutral with respect to *tonA*$^+$ under our conditions (7). Because all strains are F$^-$ and cannot undergo genetic exchange with one another, the *tonA* marker is used to monitor changes in the frequency of the *gnd* allele with which it is associated.

Chemostats. Chemostat medium consists of Davis minimal salts and one of four limiting nutrients: gluconate (0.4 g/liter); ribose + succinate (0.1 g/liter ribose and 0.3 g/liter succinate); glucose (0.4 g/liter); or fructose (0.4 g/liter). An experiment begins with inoculation of a 30-ml chemostat with a total of approximately 10^9 cells taken from exponentially growing cultures

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

Abbreviation: 6PGD, 6-phosphogluconate dehydrogenase.

Table 1. Method of strain construction

Step	P1 donor	Recipient	Selected marker	Resultant strain
1	*gndS1* (RM77C)	Δ(*zwf edd eda*) *gnd-1 his*$^-$ *rpsL* (DD715)	*his*$^+$	Δ(*zwf edd eda*) *gndS1 his*$^+$ *rpsL* (DD929)
2	DD929	DD715	*his*$^+$	Δ(*zwf edd eda*) *gndS1 his*$^+$ *rpsL* (DD953)
3	DD953	Δ(*zwf edd eda*) *gnd-1 rpsL* (DD722)	*gnd*$^+$	Δ(*zwf edd eda*) *gndS1 rpsL* (DD1095)
4	*gnd-1* (DD657)	DD1095	*eda*$^+$	*gndS1 rpsL* (DD1104)
5	Spontaneous *tonA* mutation in DD1104 yields *tonA gndS1 rpsL* (DD1113)			

of two strains, one marked with *tonA*$^-$ and the other with *tonA*$^+$, in order to give an initial frequency of *tonA*$^-$ of either 0.50 or 0.01. The rate of flow of fresh medium into the chemostat regulates the doubling time of the cells, and the concentration of limiting nutrient regulates the equilibrium population density; these are preset to correspond to a doubling time of about 2 hr and an equilibrium density of about 5×10^8 cells per ml, equilibrium being reached in about 10 hr.

Samples of cells from the overflow are collected at approximately 12-hr intervals. The total number of cells in each sample (T) is estimated by direct count of at least 350 colonies that arise on nonselective media plated with appropriate dilutions. The number of *tonA*$^-$ cells (B) is obtained similarly by plating cells with an excess of bacteriophage T5. The proportion of *tonA*$^-$ cells (R) is estimated as B/T, that of *tonA*$^+$ cells (S) as $(T - B)/T$, and the number $\ln(R/S)$ called for in the theory is estimated as $\ln[B/(T - B)]$. Data analysis consists of a linear regression of estimated $\ln(R/S)$ against time (in hours), the slope being a measure of the relative competitive ability of the two strains. The statistical significance of an observed slope is tested against the null hypothesis $s = 0$ by means of analysis of variance. Extensive computer simulations of the sampling procedure (7) have shown the estimate of $\ln(R/S)$ to be essentially unbiased as long as R is less than about 0.8. Estimates of $\ln(R/S)$ for $R > 0.8$ tend to be biased and the sampling variance is large.

RESULTS

The power of the chemostat procedure to detect selection is limited by sampling error in the estimates of $\ln(R/S)$. The theoretical effects of sampling error have been studied by means of computer simulation. Selection within a hypothetical chemostat was assumed to follow $\ln(R/S) = C + st$, and random samples corresponding in size to actual samples were generated at 12-hr intervals. Data from each hypothetical chemostat were then analyzed by regression and analysis of variance. For each specified value of s, 1024 simulations were performed. Representative results are shown in Fig. 1*A*. When alleles are in fact neutral (circles), spurious statistical significance at the 5% level is obtained 5.1% of the time, which is as expected. Note that sampling variance alone can cause individual estimates of $\ln(R/S)$ to vary between about -0.5 and $+0.5$ when the true value is 0. Triangles in Fig. 1*A* are representative simulations when $s = -0.01$. In this case, 97.1% of the regressions are significant at the 5% level, so there is only a small chance that selection of this magnitude would go undetected in any single chemostat. Representative simulations for $s = -0.005$ are illustrated by squares; 47.8% of these regressions proved to be significant at the 5% level. On the whole, because experimental chemostats are replicated at least twice, we regard selection at the level of 0.5%/hr (about 1% per generation) to be the realistic limit of resolution of the procedure.

Table 2 lists the *gnd* alleles that have been transferred to the K-12 genetic background, the designation of each original strain in the Milkman collection, the organism of origin, and the frequency of each electrophoretic type in the Milkman survey of 829 clones (1). The allele in the K-12 strain is designated *K12*; this allele codes for a 6PGD allozyme with an electrophoretic mobility like that of *W+*, which is by far the most common naturally occurring allele. Two laboratory-derived alleles have also been studied. Allele *gnd-1* is a point mutation leading to an inactive 6PGD. *Rev* arose in the third step of strain construction when a P1 lysate from a *gndS7*-bearing strain was used to transduce a *gnd-1*-bearing strain, selection being for growth on gluconate: *Rev* codes for a 6PGD having an electrophoretic mobility like that of *K12*, and it probably arose as a spontaneous reversion of *gnd-1* or as a result of intracistronic recombination between *S7* and *gnd-1*.

Experiments involving four evidently neutral alleles are illustrated in Fig. 1*B*. We adopt the convention that the strain listed first carries *tonA*$^-$ in addition to the designated *gnd* allele, and the strain listed second carries *tonA*$^+$. Lines with a positive slope indicate selection in favor of the *tonA*$^-$-bearing strain; lines with a negative slope indicate selection against the *tonA*$^-$-bearing strain. Squares in Fig. 1*B* refer to control chemostats involving competition between strains carrying *gnd-1* or *K12* in either gluconate (G) or glucose (D); selection against the *gnd-1*-bearing strain is highly significant in both cases, the estimated slopes being -0.088/hr and -0.087/hr, respectively. All of the other slopes in Fig. 1*B* are statistically nonsignificant. Thus, there is no evidence of selection involving the alleles *W+*, *F2*, *S4*, and *S7* in gluconate-limited chemostats, either in the experiments shown in Fig. 1*B* or in other experiments carried out at different initial frequencies to serve as a check against frequency-dependent selection. Although any conclusion regarding the alleles in Fig. 1*B* must be conditional on the limit of resolution of the technique, the results do support the hypothesis that the alleles in question are selectively neutral or nearly neutral with respect to one another and with respect to the alleles *K12* and *Rev*.

Fig. 1*C* pertains to an allele (*S1*) found to be detrimental in limiting gluconate. Squares refer to competition between strains bearing *S1* or *W+*, and selection against the *S1*-bearing strain occurs at the rate of about 3%/hr irrespective of whether the *S1* strain carries *tonA*$^-$ (open squares) or *tonA*$^+$ (closed squares). Closed hexagons in Fig. 1*C* indicate that the rate of selection is independent of cell density; reducing the concen-

Table 2. Alleles of *gnd* and their sources

gnd allele	Strain of origin	Organism from which strain of origin was isolated	Frequency of electrophoretic type in population (n = 829)
W+	RM43A	Human	0.713
F2	RM202F	Celebes black ape	0.036
S4	RM72B	Lowland gorilla	0.025
S8	RM200Q	Pig	0.012
S1	RM77C	Human	0.026
	RM215C	Bali steer	
S7	RM210I	Celebes black ape	0.008
K12	K-12(DD715)	Human	—
Rev	P1(S7) → DD722	(Laboratory)	—
gnd-1	DF1071 = DD657	(Laboratory)	—

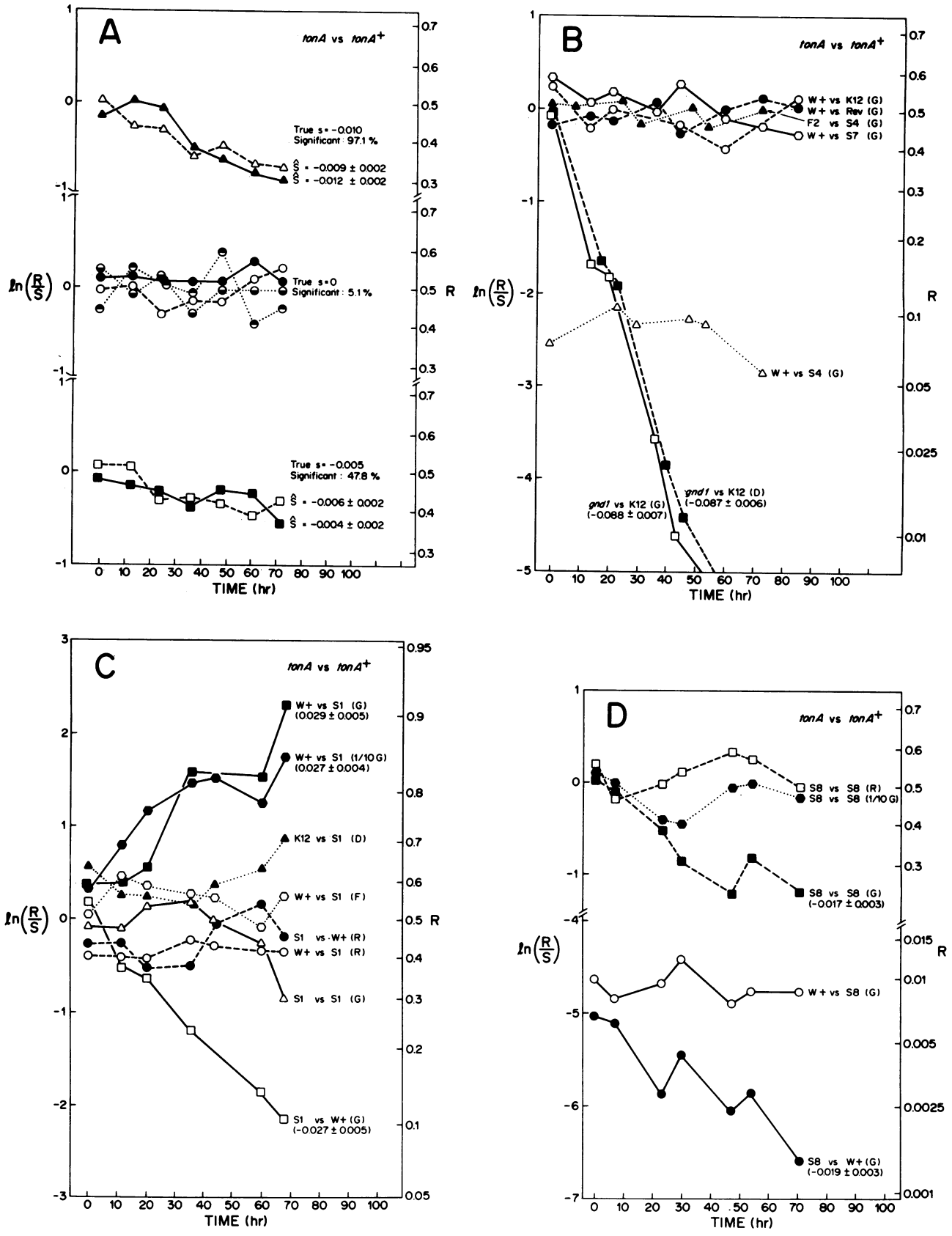

FIG. 1. (*A*) Computer simulations showing effects of sampling error for various values of *s*. Statistically significant slopes and their standard errors are indicated. Percentages refer to the proportion of 1024 simulations in which the slopes were significant at the 5% level. Parts *B–D* have the same layout. Each strain is designated by the *gnd* allele it carries, with the *tonA*-bearing strain listed first. Slopes and standard errors differing significantly from 0 are given in parentheses beneath the competing strains. Limiting nutrients are gluconate (G), glucose (D), ribose + succinate (R), or fructose (F). (*B*) Alleles *W+*, *F2*, *S4*, and *S7* are selectively equivalent in gluconate; *gnd-1* codes for an inactive 6PGD. (*C*) Allele *S1* is detrimental in gluconate but not in glucose or fructose. (*D*) Allele *S8* has a detrimental, density-dependent epistatic interaction with *tonA*.

tration of limiting gluconate 1:10 ($\frac{1}{10}$G) causes a corresponding 1:10 reduction in equilibrium cell density, yet the rate of selection against *S1* remains the same.

None of the other slopes in Fig. 1*C* is statistically significant. Open triangles are from a control experiment showing that the selection involving *S1* is not due to the associated *tonA* marker. Circles refer to control experiments in which the limiting nutrient is a mixture of ribose and succinate (R); because 6PGD is not involved in either ribose or succinate metabolism, these results imply that the selection observed with gluconate limitation is due to the *S1* allele itself and not due to some unrecognized mutation in the genetic background. Closed triangles and open hexagons in Fig. 1*C* refer to chemostats limited for glucose (D) or fructose (F); in contrast to the situation with *gnd-1*, in which selection under gluconate limitation is of about the same magnitude as under glucose limitation (Fig. 1*B*), the selection associated with *S1* observed in gluconate-limited chemostats disappears in chemostats limited for glucose or fructose.

The experiments in Fig. 1*C* involve the *S1* allele from Milkman strain RM77C, which was isolated from a human fecal sample. Because this particular allele could be atypical, similar experiments were carried out with strains bearing the *S1* allele from strain RM215C, isolated from a Bali steer. Gluconate-limited chemostats containing *gnd W+ tonA*$^-$ versus *gnd S1 tonA*$^+$ yielded highly significant slopes of 0.036 ± 0.004 and 0.041 ± 0.002 in two replicates, and a similar chemostat involving *gnd S4 tonA*$^-$ versus *gnd S1 tonA*$^+$ produced a slope of 0.032 ± 0.003, all values being in reasonable accord with those reported in Fig. 1*C*. The *S1* allele thus seems to be detrimental in gluconate-limited chemostats irrespective of its strain of origin.

Fig. 1*D* illustrates the characteristics of selection discovered with another allele (*S8*). The selection has two unusual features. First, it involves an epistatic interaction with another allele in the genetic background, in particular with *tonA*. Closed circles indicate that the *S8–tonA*$^-$ combination is disfavored at the rate of about 2%/hr in gluconate-limited chemostats, yet the open circles indicate that the *S8–tonA*$^+$ combination is selectively neutral. Direct competition between *gnd S8 tonA*$^-$ and *gnd S8 tonA*$^+$ is shown by the closed squares; selection again disfavors the *S8–tonA*$^-$ combination at the rate of about 2%/hr. Although the metabolic basis of the *S8–tonA*$^-$ interaction is not known, the selection evidently involves gluconate utilization, because it is not observed in chemostats limited for ribose + succinate (open squares). The second unusual feature of selection involving *S8* is illustrated in Fig. 1*D* by the closed hexagons; the selection is density dependent. When the equilibrium cell density is reduced 1:10 by a 1:10 reduction in the concentration of limiting gluconate ($\frac{1}{10}$G), the selection involving *S8–tonA* is no longer observed.

To determine whether the *S8–tonA* interaction might be due to an atypical *tonA* allele, two additional strains were obtained by selection of spontaneous T5-resistance *tonA* mutations in the *gnd S8 tonA*$^+$ strain. In gluconate-limited chemostats involving *gnd S8 tonA*$^-$ versus *gnd S8 tonA*$^+$, the first of these new strains yielded a slope of −0.018 ± 0.001, the second a slope of −0.016 ± 0.003, both values very close to the value − 0.017 ± 0.003 recorded in Fig. 1*D*.

DISCUSSION

The most conservative interpretation of the finding of no selection involving *W+*, *F2*, *S4*, and *S7* in gluconate-limited chemostats is that the magnitude of the selective differences falls below the limit of resolution of the technique—approximately 0.5%/hr or 1% per generation. On the other hand, gluconate was chosen as the limiting nutrient because the functional effectiveness of 6PGD is a principal determinant of growth on gluconate. Because gluconate limitation is probably a rare occurrence in natural environments, any selection found with gluconate limitation is likely to overestimate the true amount of selection that occurs in nature. In a broad sense, of course, the ability to compete for limiting nutrient is only one aspect of bacterial fitness. Other components include ability to colonize habitats, persistence in the face of starvation, ability to compete against other species, and so on. With regard to the allozymes of 6PGD, however, differences in any other of these fitness components must ultimately be traceable to differences in utilization of substrate by the allozymes, because 6PGD plays a specialized and highly specific role in metabolism. Because our experiments involving nutrient limitation have been designed to maximize selection based on substrate utilization, the simplest interpretation regarding the alleles found to be neutral in gluconate-limited chemostats is that the alleles are also neutral relative to fitness in the broad sense. It therefore seems justified to regard the alleles in question as being neutral or nearly neutral in the sense that the selective differences among them are sufficiently small that their changes in frequency are dominated by random genetic drift. In *E. coli*, a principal source of random genetic drift is frequent extinction and recolonization of local populations, which greatly reduces the effective population size and thus accelerates random genetic drift (8).

How the results regarding *S1* and *S8* relate to natural populations is unclear. The alleles are undoubtedly associated with selection under gluconate limitation, but the selection is complex and the prerequisites for its expression may not often prevail in natural populations. The *S1* allele is selectively neutral in chemostats limited by glucose or fructose, and the density-dependent selection involving *S8* is due to an epistatic interaction with *tonA*. These complex characteristics imply that the magnitude of selection in natural populations will depend on the particular environment, on other alleles in the genetic background, on population density, and perhaps on other factors not yet identified. Any or all of the prerequisites for selection may on occasion occur in nature, in which case alleles that are otherwise nearly neutral may become subject to selection. An accepted terminology for this situation is unfortunately not available, but the data suggest that otherwise neutral or nearly neutral alleles may have a latent potential for selection that can be realized under the appropriate conditions.

One additional *gnd* allele warrants mention because it seems to represent a harmful allele maintained by mutation–selection balance. The allele is found in strain BLQQ and codes for an enzymatically inactive 6PGD. Although the allele has not been transferred into a K-12 background for detailed study, it seems likely that its selective effects would be similar to those of *gnd-1*, which is highly detrimental in chemostats limited for either gluconate or glucose (Fig. 1*B*). The frequency of the *gnd* null allele in natural clones is consistent with its being a harmful allele maintained by recurrent mutation: no *gnd* nulls were found in Milkman's study of 829 clones (1), but the one in BLQQ was found in Selander and Levin's study of 19 additional clones (2), giving an overall frequency of the null as 1/848.

On the whole, our results with *gnd* support one version of the neutral theory of molecular polymorphism and fail to support another. The view supported is that the expressed selective effects of allozymes are typically so small that allele-frequency changes are largely determined by random genetic drift (4). The view not supported is the simplistic caricature of the neutral theory which holds that allozymes are functionally irrelevant to the organism. Allozymes do indeed have functional differences; the corresponding alleles thereby have a potential for selection that can become expressed under the appropriate conditions of

environment or genetic background. Potential for selection is exemplified in the present study by the alleles *S1* and *S8*, but we have found previously that even *F2* and *S4* are nonneutral in gluconate-limited chemostats when the genetic background contains an *edd*$^-$ mutation that cuts off the alternative metabolic route for 6-phosphogluconate (7). The allozymes of 6PGD are functionally diverse and selection, when it occurs, has complex characteristics. Thus, the overall picture that emerges for *gnd* is one of near neutrality superimposed upon unexpressed but potentially important functional differences among alleles. Whether this picture will apply to other polymorphic loci in *E. coli* is an issue yet to be resolved.

We acknowledge the expert technical assistance of Steven Newton and thank B. Bachmann, B. Levin, and R. Milkman for providing strains. This work was supported by Grant GM 24886 from the National Institutes of Health.

1. Milkman, R. (1973) *Science* **182**, 1024–1026.
2. Selander, R. K. & Levin, B. R. (1980) *Science* **210**, 545–547.
3. Kimura, M. (1968) *Nature (London)* **217**, 624–626.
4. Kimura, M. (1977) *Scientia* **112**, 687–707.
5. Dykhuizen, D. E. & Hartl, D. L. (1981) *Evolution* **35**, 581–594.
6. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
7. Dykhuizen, D. E. & Hartl, D. L. (1980) *Genetics* **96**, 801–817.
8. Maruyama, T. & Kimura, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6710–6714.