J Mol Evol (1996) 42:79–84

JOURNAL OF MOLECULAR EVOLUTION

Low Synonymous Site Variation at the *lacY* Locus in *Escherichia coli* Suggests the Action of Positive Selection

Robert R. Wagner, Margaret A. Riley

Department of Biology, Osborn Memorial Laboratories, Yale University, 165 Prospect Street, New Haven, CT 06511, USA

Received: 1 April 1995 / Accepted: 15 September 1995

Abstract. We have determined the nucleotide sequences of seven *lacY* alleles isolated from natural isolates of *Escherichia coli.* Nucleotide heterozygosity estimates for this locus were compared to those obtained from previous studies of intraspecific variation at chromosomal loci, revealing that *lacY* has unusually low synonymous site variation. The average pairwise heterozygosity of synonymous sites ($Ks = 0.0112 +/- 0.0100$) is the second lowest reported and the lowest for loci that have an equivalent level of nonsynonymous variation. We consider several hypotheses to explain how different forces in evolution could act to create the observed pattern of polymorphism, including selection for translational efficiency and positive selection. Our analysis most strongly supports the hypothesis that positive selection has acted on the *lacY* locus in *E. coli.*

Key words: *LacY* — Lactose permease — Evolution — Reduced synonymous polymorphism — Positive selection — *E. coli*

Introduction

The description of polymorphism in the genetic loci of *Escherichia coli* has added much to our knowledge of how the forces of mutation, genetic drift, and selection have influenced organismal evolution. Studies of nucleotide polymorphism in alleles of chromosomally encoded loci *phoA* (Dubose et al. 1988), *trp* (Milkman and Crawford 1983), and *gnd* (Bisercic et al. 1991) have shown the degree to which point mutation and recombination have affected the divergence of genes, while observation of nucleotide polymorphism at *gapA* and *pabB* (Guttman and Dykhuizen 1995) suggests the occurrence of selective sweeps which homogenize large expanses of the chromosome as a consequence of genetic hitchhiking in the absence of recombination. Using polymorphism data, these studies and others (Hall and Sharp 1992) have allowed investigators to infer the genealogical relationships between subpopulations within a species and to reconstruct the evolutionary histories of individual loci.

A detailed description of naturally occurring polymorphism in alleles of a single genetic locus and of the effects of that variation on selective advantage or fitness is absent from the literature describing populations of enteric bacteria. While it is clear that phenotypic variation exists and that selection can act on this variation to shape the composition of populations, there are no studies that describe polymorphism found at loci whose variant alleles are known to confer different levels of fitness. Previous attempts to establish a cause and effect relationship between enzyme polymorphism and variation in physiology and reproductive success of eukaryotes show correlations between the prevalence of particular isozymes and environmental conditions in which the examined isozymes are purported to confer a physiological advantage (reviewed by Koehn et al. 1983; Gillespie 1991). An inference from these studies is that kinetic variation in the studied enzymes is subject to selection and that the expression of certain isozymes in some en-

Correspondence to: M.A. Riley

vironments is beneficial for survival and enhances organismal fitness.

The ability of such studies to link polymorphism to variation in fitness is weakened because observations of kinetic variation were made in vitro, eliminating epistatic interactions which have the potential to mask the effects of kinetic variation on physiology and phenotype. Many of these studies are additionally limited by the use of experimental systems that utilize organisms that are not amenable to the comparative assessment of fitness among more than a few hundred or thousand individuals over a time period of sufficient duration to allow subtle, evolutionarily significant differences in fitness to visibly affect population composition.

There is one set of studies that reveals the relevance of polymorphism at a single locus to organismal fitness (Dykhuizen et al. 1987; Dean 1989). The investigators performed empirical and theoretical work showing that variation in *E. coli* lactose permease, encoded by the *lacY* locus, has a more significant effect on fitness than variation in *lacZ*-encoded beta-galactosidase. Using isogenic strains that differed only at the *lac* operon, the investigators demonstrated that variation in *lac* genes produced significant fitness differences in a lactose-limited chemostat. To determine the degree to which each locus in the region affected fitness, metabolic control theory (Kascer and Burns 1973; Kascer and Burns 1981) was used to show that the product of *lacY* has a greater ability to limit the catabolism of lactose than does the product of the *lacZ* gene. This finding suggests that existing variation in the *lacY* gene is relevant to fitness and that the pattern of polymorphism at this locus is likely to be characteristic of a locus under selection.

This study presents observations of polymorphism at the *lacY* locus in natural populations of *E. coli.* We present the DNA sequences of seven *lacY* alleles, four of which have documented effects on fitness in the lactose-limited chemostat (Dean 1989), and an analysis of the patterns of nucleotide and amino acid variation. The analysis of these patterns is then used to test hypotheses that suggest the kind of selective constraints operating at the *lacY* locus. The data presented here add to the growing database of within-species variation in bacteria and serve as a foundation for further studies designed to determine the effects of polymorphism on fitness.

Materials and Methods

The *Escherichia coli* strains used as a source of *lacY* DNA are from the ECOR collection (Ochmann and Selander 1984) or contained a *lacY* allele that had been transduced from an ECOR strain into a common genetic background. The strains, TD 1, TD 9, TD 10, and TD 13 (Dean 1989), contained the *lacY* gene of strains CSH 64 (Miller 1972), ECOR 4, ECOR 16, and ECOR 70, respectively. The four TD strains were chosen because the effects of variation at *lacY* on the fitness of these strains has been documented. Three additional strains, ECOR 1, ECOR 50, and ECOR 71, were chosen as source strains to increase diversity of the *lacY* allele sample.

The ECOR sources of *lacY* alleles described in this study include members of three of the five major strain groups described by Selander et al. (1987). It is expected that this sample has a level of diversity comparable to that of previous studies at other loci (Table 3) which, with the exception of studies at *trp,* used source strains from the same strain groups. Our study includes *lacY* alleles from four of the source strains used in the study of the *phoA* locus (Dubose et al., 1988) and from three used in the study of *gapA* (Gutmann and Dykhuizen, 1994).

ECOR and TD strains were grown overnight in LB liquid medium, and these cultures were used as a source of genomic DNA for polymerase chain reaction amplification of *lacY.* Regions of the *E. coli* chromosome which contain *lacY* were amplified via the polymerase chain reaction (Innis et al. 1990) using between 100 and 400 ng genomic DNA in a 100-μl total volume containing 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 3 mM $MgCl_2$, 0.001% gelatin, 200 μM each dNTP, 1 unit Amplitaq polymerase, and 40 nM each of two primers that flank both ends of the *lacY* coding sequence (Buchel et al. 1980). Primer lacY1B (5′-CACATGGCTGAATATCGAC-3′) has identity with the coding strand of *lacZ* at nucleotide position 4229–4248 of the GeneBank file ECOLAC (accession #K01483), which contains a sequence of the *lac* region. Primer lacY2B (5′-GGAATTCCGCATGTTCAATGCGATCAC-3′) is complementary to the coding strand at nucleotides 5736–5718 and includes an additional eight nucleotides at its 5′ end which encode an *Eco*RI restriction site. Amplification was carried out in a Hybaid OmniGene thermal cycler running in the ''tube control'' mode for 35 amplification cycles with a 94°C denaturation step for 1 min, a 55°C anneal step for 30 s, and a 72°C extension step for 1 min. Cycling was preceded by a 5-min incubation at 90°C and followed by a 3-min final extension at 72°C. The annealing temperature was the same for all reactions except those containing genomic DNA from strains ECOR 50 and ECOR 71; for these reactions, an annealing temperature of 40°C was necessary to produce a sufficient amount of product for subsequent cloning.

PCR products were ligated to the pcrII vector following the TA cloning kit protocol (Stratagene). Plasmids were obtained from transformed strains using the alkaline lysis method (Maniatas et al. 1982). DNA sequence was determined using the dideoxyribonucleic acid chain termination method with denatured plasmid templates as prescribed in the Sequenase version 2.0 Sequencing Kit (United States Biochemical). Primers used in sequencing are as follows (name/sequence identity with ECOLAC bp#/sequence): lacY1B, lacY3/4629–4653/5′-ATTACCTGCTGTGGATTATTACCG-3′, lacY4/4913–4933/5′-TGTTTTCTGGCTGGGCTCTG-3′, lacY5/5108–5131/5′-CTGCACCTACGATGTTTTTGACC-3′, and lacY7/5331–5351/5′-TTCGCCACCTCAGCGCTGGA-3′. The products of the sequencing reaction were resolved on a denaturing 6% acrylamide gel.

Results

Twelve hundred fifty-four base pairs of the coding region of *lacY* (Buchel et al. 1980) were sequenced in each of the seven cloned alleles. The location and type of polymorphism observed in DNA sequences are presented in Table 1, where the published sequence is used as the reference for all comparisons. The location of nonsynonymous polymorphic sites in the putative structure of the permease protein (Kaback et al. 1993) is indicated in Table 2. Table 1 shows that each allele has a unique combination of polymorphic sites. The synonymous substitutions of cytosine for guanosine at position 370 and guanosine for cytosine at position 371 are ubiquitous in each allele, and the nonsynonymous substitu-

Table 1. Polymorphic sites in *lacY* alleles[a]

Source	Nucleotide position in *lacY*: 150	241	291	303	342	370	371	409	502	503	560	582	602	673	832	867	1071	1073	1078	1145	1168	1212	1221	1224	1229	1250	1254
Published:	c	a	a	c	a	g	c	a	t	t	t	t	t	t	t	a	t	a	c	a	c	c	c	g	g	c	a
TD 1			g			c	**G**								**C**											**A**	
TD 9						c	**G**	**G**																			
TD 10						c	**G**				**C**		**C**	c		g											
TD 13						c	**G**																				
ECOR 1		**T**				c	**G**													**G**							
ECOR 50	t				t	c	**G**					c							t		t	t	t	a	**A**		
ECOR 71				t		c	**G**		**C**	**C**							c	**G**							**A**		g

[a] Reference sequence (Buchel et al. 1980) is in *italics.* Position 1 is adenine of the start codon for *lacY* and corresponds to position 4410 of the GeneBank sequence for the *lac* operon (accession # K01483). Nonsynonymous substitutions are in uppercase boldface (**G**); synonymous substitutions are in lowercase (g)

Table 2. Naturally occurring amino acid variation in lactose permease

Residue position	Domain[a]	Strain: Published	TD 1	TD 9	TD 10	TD 13	ECOR 1	ECOR 50	ECOR 71
81	Helix III	Thr					Ser		
124	Helix IV	Ala	Arg	Arg	Arg	Arg	Arg	Arg	Arg
137	Loop 5	Asn		Asp					
168	Helix VI	Phe							Pro
187	Helix VII	Ala			Asp				
201	Loop 7	Val			Ala				
278	Helix VIII	Phe	Leu						
358	Helix XI	Lys							Arg
382	Helix XII	Tyr					Cys		
410	Loop 13	Arg						His	His
417	Loop 13	Ala	Asp						

[a] Location of variable sites within secondary-structure elements is inferred from Kaback et al. (1993)

tion of an adenine for a guanine at position 1229 is shared between strains ECOR 50 and ECOR 71. No other polymorphic sites are shared between alleles.

A statistical evaluation of the location of polymorphic sites was performed to determine whether polymorphic sites are clustered within the gene and the encoded permease protein. The polymorphic sites listed in Table 1 were mapped onto a single *lacY* sequence, and a runs test (Sokal and Rohlf 1981) was applied to the pooled data. The standard normal deviate for the clustering of all polymorphic sites in *lacY* is −1.91, which suggests a departure from random distribution. It is likely that this result is significantly influenced by the clustering at positions 370 and 371, and at positions 502 and 503. The test was repeated using either total synonymous sites, total nonsynonymous sites, or amino acid substitutions occurring in the permease polypeptide. The tests of synonymous and amino acid substitutions produced standard normal deviate values of 0.89 and 0.55, respectively, suggesting that variation of silent sites and at amino acids within the permease is random. The distribution of nonsynonymous sites departed from random expectation. The apparent clustering of nonsynonymous substitution may be a consequence of the clustered pair of substitutions at positions 502 and 503 in strain ECOR 71 (Table 1) and is at odds with the results of runs test on the distribution of amino acid polymorphism.

An examination of the distribution of amino acid polymorphism among the structural domains of the lactose permease shows that variable residues are equally partitioned among secondary-structure elements. Table 2 shows that 4 of 11, or 36%, of polymorphic amino acid positions are located in helix spanning loops which are comprised of 37% of the amino acids in the permease (Kaback et al. 1993). At two of these positions, amino acids 137 and 417 in strains TD 9 and TD 1, respectively, the amino acids at the polymorphic sites have a charge character different from that of the amino acid encoded by the published sequence. Of the six polymorphic positions encoding a residue in purported transhelical do-

Table 3. Polymorphism in *lacY* DNA sequences obtained from *E. coli*[a]

	lacY vs	*celc*	*crr*	*gutb*	*phoa*	*trpB*	*trpc*	*gapA*
Total sites	1254	351	510	372	1416	1194	1359	937
# total poly	27							
% poly	2.15							
Kt	0.0056 ± 0.0030	0.012	0.011	0.014	0.02	0.011	0.022	0.0028
# syn sites	322							
# syn poly	15							
% syn poly	4.66							
Ks	0.0115 ± 0.0100	0.049	0.051	0.044	0.071	0.049	0.076	0.0028
# nonsyn sites	932							
# ns poly	12							
% ns poly	1.29							
Kns	0.0036 ± 0.0010	0.0014	0	0.0057	0.0036	0	0.0038	0
Kns/Ks	0.313	0.03	0	0.13	0.05	0	0.05	0
Position	8	38	52	58	9	28	28	35
CAI	0.466	0.34	0.59	0.35	0.33	0.41	0.311	0.826

[a] Abbreviations: syn, synonymous: nonsyn, nonsynonymous; *Ks*, average basepair heterozygosity at synonymous sites; *Kns*, average basepair heterozygosity at nonsynonymous sites. Data for loci *crr, gutB, celC, trpB,* and *trpC* are from Hall and Sharp (1992), for *phoA* from Dubose et al. (1988), for *gapA* from Gutmann and Dykhuizen (1995). Chromosomal positions are in minutes according to the *E. coli* K-12 map of Bachmann (1990)

mains only one, position 187 from strain TD 10, has a difference that includes an altered charge character.

The results from additional analysis of the observed polymorphism are shown in Table 3, which presents average heterozygosity estimates (Li et al. 1985) for *lacY* and for several other loci whose *K* values were calculated in previous studies. The average nucleotide diversity for all positions (*Kt*) in *lacY* is the second smallest figure for a chromosomally encoded locus in *E. coli;* only *gapA* has a smaller average number of differences between alleles. Inspection of estimates for *Kns* and *Ks* reveals the cause of the low *Kt* value. The average difference between synonymous sites, *Ks,* for *lacY* is two- to sevenfold less than for other loci with the exception of *gapA,* while the average difference between nonsynonymous sites, *Kns,* is within the range defined by the values determined for *celC* and *gutB.*

The codon adaptation index (CAI) (Sharp and Li 1987) value for *lacY* and additional loci is presented at the bottom of Table 3. The CAI value obtained for *lacY* is 0.466, typical of genes that are translated at an intermediate expression level. A comparison of the Relative Synonymous Codon Usage (RSCU) values (Sharp and Li 1986) obtained for individual codons in *lacY* to those of genes with high or low expression level showed that codons for more than half of the amino acids encoded by multiple codons have RSCU values close to genes expressed at low levels (data not shown).

The previous tests do not show a correlation between codon usage and the level of synonymous polymorphism at *lacY,* so the relationship between the chromosomal position of *lacY* and synonymous polymorphism was investigated. Other studies (Sharp 1991; Sharp et al. 1987) indicate locus position relative to *oriC* is correlated with the value of *Ks* when the locus under investigation has a CAI less than 0.500. Genes further from *oriC* tend to have an increased *Ks* value. This value for *lacY* is much lower than for loci that are closer to *oriC,* as might be expected, but is equally less than that of loci located farther away.

Discussion

Analysis of the DNA sequences of seven *lacY* alleles obtained from naturally occurring isolates of *E. coli* shows that *lacY* has the second lowest level of synonymous site variation of any *E. coli* gene examined to date. Only *gapA,* a locus that appears to have been purged of diversity during a selective sweep, has less synonymous site diversity (Guttman and Dykhuizen 1995). A comparison of *Ks* values presented in Table 3 shows that the average number of differences at synonymous sites between any pair of *lacY* alleles is nearly sevenfold lower than is seen in *phoA* and *trpC,* which have an approximately equal level of nonsynonymous (*Kns*) variation. These two observations suggest that forces of evolution have acted on *lacY* in a way that either limits evolution at synonymous sites or increases variation at nonsynonymous sites.

Previous studies have shown that genes with high expression levels utilize a limited set of codons that are purported to be optimized for rapid translation (Sharp and Li 1986). The CAI and RSCU estimates for *lacY* both suggest that the locus is not subject to strong selection for translational efficiency and hence is not expected to exhibit an unusual degree of constraint at synonymous sites. To determine whether the low synonymous variation in *lacY* is typical of a gene with a purported intermediate expression level, we compare the CAI values and *Ks* estimates between loci in Table 3. The *Ks* value of *lacY* is larger than that for *gapA,* which is as expected

because of the high expression level indicated for *gapA* by its CAI. The *Ks* value of *lacY* is severalfold smaller than that for *trpB*, whose CAI is closer to that of *lacY*. This suggests that a greater number of silent sites in *lacY* could be polymorphic without effect on the level of expression. If selective constraint is maintaining low variation at silent sites in *lacY*, it appears to be acting for a reason other than the maintainence of codons at a frequency in proportion to tRNA abundance.

The relevance of chromosomal placement in explaining the low level of *Ks* variation at *lacY* is also unresolved. Previous work (Sharp et al. 1989) suggests a tendency for the between-species *Ks* values of chromosomally encoded loci to increase as their distance from *oriC* becomes larger, but it is difficult to determine whether the limited synonymous variation at *lacY* is as expected for loci located at minute 8 (Bachmann 1990). The difficulty is due to the small number of intraspecies estimates of *Ks* that are available and the absence of diversity data for *lacY* in *Salmonella typhimurium* which would allow calculation of an interspecies *Ks* value that could be compared to those previously reported (Sharp 1991) in studies of the correlation between locus position and *Ks*. The within-species estimates of *Ks* (Table 3) for *E. coli* do not show a clear trend in their relationship to distance from *oriC*. The loci *gutB*, *crr*, *trpb*, and *trpC* are located at approximately equal distances from *oriC* and are expected to have roughly equivalent *Ks* values. This is not the case.

A previous study (Dean 1989) of the phenotypic effects of the alleles from strains TD 1, TD 9, TD 10, and TD 13 suggests another type of selection which may shape the evolution of the *lacY* allele pool in a way that decreases synonymous site variation. This study clearly illustrates the ability of slight variation in permease activity to affect strain fitness and indicates that the polymorphisms presented in Tables 1 and 2 may have significant effects on selection coefficients. Thus it is possible that the pattern of polymorphism observed in *lacY* is due to the effects of positive selection (Tanaka and Nei 1989; Ohta 1992) acting to increase the frequency of specific alleles within the *E. coli* population. The high *Kns*/*Ks* ratio seen in Table 3 is compatible with the effects of such selective sweeps. Synonymous site variation remains low because mutation at nonsynonymous sites creates fitness-enhancing alleles that are recurrently swept through the population before mutation can saturate synonymous sites.

Since the analysis of codon usage bias and chromosomal position does not imply unusually high functional constraint on synonymous sites, the hypothesis that the low divergence at these sites is due to positive selection appears more plausible. The acceptance of this hypothesis is further justified by the demonstration that the lactose permease has a high metabolic control coefficient relative to other enzymatically catalyzed steps in the catabolism of lactose (Dykhuizen et al. 1987; Dean 1989), which suggests that amino acid variation will be subject to selection when lactose is a limiting resource. Alleles with a nucleotide variation at nonsynonymous sites that leads to only slight kinetic variation could quickly dominate in a population of strains with permeases that transport less lactose.

The type and location of amino acid substitutions that are a potential cause of interstrain fitness differences in nature are presented in Table 2. Mapping the position of each nonsynonymous site onto a protein sequence secondary-structure model (Kaback et al. 1993) of the permease reveals that two-thirds of the polymorphic amino acid residues are located in the 12 membrane-spanning helical domains. The remainder are located in trans-helix loops located on the cytoplasmic side of the lipid bilayer, connecting helix IV to V and VI to VII. Our statistical analysis of the distribution of amino acid substitution suggests a random distribution and suggests that no region of the lactose permease is under unusual constraint or is unusually variable.

Polymorphism in the transmembrane domains has the ability to affect substrate specificity, proton symport, and the formation of salt bridges that define helix–helix interactions within the lipid bilayer. The polymorphism at position 358 in strain ECOR 71 may affect permease structure by altering salt bridge formation (Kaback et al. 1993; Jung et al. 1993; Sahin-Toth and Kaback 1993), as the lysine found at this position has been implicated in the formation of a bridge with Asp237 in helix VII. The substitution of lysine with arginine at residue 358 retains a positively charged side chain which could maintain a stable association with the basic side chain of Asp237, although a change in the shape of the side-chain terminus may affect the strength of the putative bridge and hence the behavior of the molecule.

The remainder of the variable residues in the transmembrane domains are positions whose effects on lactose transport have not been described. Two of these positions are near charged residues, one of which may participate in salt bridging. The substitution of Asp188 with Ala in the permease of strain TD 10 replaces a charged polar side chain with an apolar side chain at a position adjacent to positively charged Lys189. This replacement may change the charge character of the protein near the interface between the bilayer and the cytosol. The substitution of Leu for Phe at position 277 in strain TD 1 could affect the salt bridge formed by Glu 269 (Franco and Brooker 1994), producing observable effects on lactose uptake due to changes in the structural integrity of the permease at helix VIII or X.

One weakness of the hypothesis of positive selection acting at the *lacY* locus is the absence of reduced levels of diversity in neighboring *phoA* (Dubose et al. 1988). A reduction of diversity at this locus is expected due to the effects of hitchhiking by loci near *lacY* after a selective sweep has occurred. Despite the absence of evidence for a selective sweep in the *lac-pho* region, we consider the

possibility that our observations have revealed the effects of positive selection at *lacY* and that recombination at *phoA* followed the most recent selective sweep. Since the distance between *lacY* and *phoA* is as large or larger than a fragment of DNA likely to be introduced by phage P1-mediated transduction, it is possible that recombination at *phoA* did not include any of the *lac* region. This would leave the pattern of polymorphism that remained following positive selection at *lacY* intact.

While it is not certain that positive selection has taken place at *lacY*, a reconsideration of the effects of naturally occurring polymorphism at the locus on strain fitness indicates the potential for rapid shifts in population structure and clone abundance as a consequence of changes in the ability to utilize a limiting nutrient (Dykhuizen et al. 1987; Dean 1989). As a result of its role as a metabolic checkpoint with large control on the entrance of lactose into metabolism, the lactose permease is potentially subject to significant selective pressure which acts rapidly to increase the frequency of permease variants that confer increased fitness when lactose is available.

Acknowledgments. The authors wish to thank Anthony Dean and Daniel Dykhuizen for kindly providing the TD strains used in this study and Robert Dorit, Junhyong Kim, Michael Feldgarden, and Jennifer Wernegreen for their critical reviews of the manuscript during its preparation. This work was supported by an NSF Young Investigator Award (DEB-9458247), by an NIH First Award (GM 47471), by a grant from the General Reinsurance Co. to M.A.R., and by a NIH Predoctoral Training Program grant (GM07499) to R.R.W.

References

Bachmann BJ (1990) Linkage map of *Escherichia coli* K-12, edition 8. Microbiol Rev 54(2):130–197

Bisercic M, Feutrier JY, Reeves PR (1991) Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli:* evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J Bacteriol 173:3894–3900

Buchel DE, Gronenborn B, Muller-Hill B (1980) Sequence of the lactose permease gene. Nature 283:541–545

Dean AM (1989) Selection and neutrality in lactose operons of *Escherichia coli.* Genetics 123:441–454

Dykhuizen DE, Dean AM, Hartl DL (1987) Metabolic flux and fitness. Genetics 115:25–31

Dubose R, Dykhuizen DE, Hartl D (1988) Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli.* Proc Natl Acad Sci USA 85:7036–7040

Franco PJ, Brooker RJ (1994) Functional roles of Glu-269 and Glu-325 within the lactose permease of *Escherichia coli.* J Biol Chem 269: 7379–7386

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, New York, pp 1–40

Guttman DS, Dykhuizen D (1994) Detecting selective sweeps in naturally occurring *Escherichia coli.* Genetics 138:993–1003

Hall BG, Sharp PM (1992) Molecular population genetics of *Escherichia coli:* DNA sequence diversity at the *celc, crr,* and *gutB* loci of natural isolates. Mol Biol Evol 9:654–665

Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990) PCR Protocols: a guide to methods and applications. Academic Press, San Diego, pp 3–20

Jung K, Jung H, Wu J, Prive GG, Kaback HR (1993) Use of site directed fluorescence labeling to study proximity relationships in the lactose permease of *Escherichia coli.* Biochemistry 32:12273–12278

Kaback HR, Jung K, Jung H, Wu J, Prive G, Zen K (1993) What's new with lactose permease. J Bioenerget Biomembr 25(6):627–636

Kascer H, Burns JA (1973) The control of flux. Symp Soc Exp Biol 27:65–104

Kascer H, Burns JA (1981) The molecular basis of dominance. Genetics 97:639–666

Koehn RK, Zera AJ, Hall JG (1983) Enzyme polymorphism and natural selection. In: Nei M, Koehn RC (eds) Evolution of genes and proteins. Sinauer Associates, Sunderland, pp 115–136

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 1.38

Milkman R, Crawford IP (1983) Clustered third base substitutions among wild strains of *Escherichia coli.* Science 221:378–379

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p 21

Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690–691

Ohta T (1992) A statistical examination of hypervariability in complementarity-determining regions of immunoglobulins. Mol Phylogenet Evol 1(4):305–311

Sahin-Toth MJ, Kaback HR (1993) Properties of interacting aspartic acid and lysine residues in the lactose permease of *Escherichia coli.* Biochemistry 32(38):10027–10035

Selander RK, Caugant DM, Whittam TS (1987) Genetic structure and variation in natural populations of *Escherichia coli.* In: Neidhardt FC (ed in chief) *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biology. American Society for Microbiology, Washington, DC

Sharp PM (1991) Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium:* codon usage, map position, and concerted evolution. J Mol Evol 33:23–33

Sharp PM, Li W (1986) An evolutionary perspective on synonomous codon usage in unicellular organisms. J Mol Evol 24:28–38

Sharp PM, Li W (1987) The codon adaptation index—a measure of directional synonomous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295

Sharp PM, Shields DC, Wolfe KH, Li W (1989) Chromosomal location and evolutionary rate variation in enterobacterial genes. Science 246:808–810

Sokal RR, Rohlf FJ (1981) Biometry, 2nd ed. WH Freeman and Company, New York, p 783

Tanaka T, Nei M (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Mol Biol Evol 6:447–459