Balancing Selection on Electrophoretic Variation of Phosphoglucose Isomerase in Two Species of Field Cricket: *Gryllus veletis* and *G. pennsylvanicus*

Laura A. Katz and Richard G. Harrison

Section of Ecology and Systematics, Corson Hall, Cornell University, Ithaca, New York 14853

Manuscript received December 6, 1996

Accepted for publication May 29, 1997

ABSTRACT

Two species of crickets, *Gryllus veletis* and *G. pennsylvanicus*, share six electrophoretic mobility classes for the enzyme phosphoglucose isomerase (PGI), despite evidence from other genetic markers that the two species are not closely related within eastern North American field crickets. Moreover, the frequencies of the two most common PGI electrophoretic classes (PGI-100 and PGI-65) covary in sympatric populations of these species in the eastern United States, suggesting that PGI may be subject to trans-specific balancing selection. To determine the molecular basis of the electrophoretic variation, we characterized the DNA sequence of the *Pgi* gene from 29 crickets (15 *G. veletis* and 14 *G. pennsylvanicus*). Amino acid substitutions that distinguish the electrophoretic classes are not the same in the two species, and there is no evidence that specific replacement substitutions represent trans-specific polymorphism. In particular, the amino acids that diagnose the PGI-65 allele relative to the PGI-100 allele differ both between *G. veletis* and *G. pennsylvanicus* and within *G. pennsylvanicus*. The heterogeneity among electrophoretic classes that covary in sympatric populations coupled with analysis of patterns of nucleotide variation suggest that *Pgi* is not evolving neutrally. Instead, the data are consistent with balancing selection operating on an emergent property of the PGI protein.

WITH the advent of protein gel electrophoresis, evolutionary geneticists gained access to an important tool for documenting patterns of variation in natural populations (e.g., HARRIS 1966; LEWONTIN and HUBBY 1966). Data from electrophoretic surveys quickly revealed that populations harbor considerable amounts of variation at protein-coding loci. These observations initiated a debate on whether such variation is neutral or maintained by balancing selection (LEWONTIN 1974). Balancing selection is defined here as any sort of selection that maintains variation within populations or species (e.g., heterozygote advantage, frequency dependent selection, spatially varying selection).

Observations of patterns of variation (e.g., heterozygote excess, clinal variation, or parallel variation) originally provided indirect evidence for balancing selection. Far more powerful evidence supporting a role for selection in maintaining enzyme polymorphisms came from demonstration of kinetic and/or thermostability differences among different genotypes and the ability to correctly predict differences in organism performance (see reviews in GILLESPIE 1991; POWERS *et al.* 1991; WATT 1994). For example, in *Colias* butterflies biochemical properties of different phosphoglucose isomerase (PGI) genotypes predict differences in survivorship, flight behavior, male mating success and female fecundity (WATT 1983, 1992, 1994; WATT *et al.*

1983, 1985). *In vitro* and *in vivo* studies of enzyme function (e.g., PLACE and POWERS 1979; WATT 1983; LABATE and EANES 1992) and field studies of genotype-dependent performance (e.g., DIMICHELE and POWERS 1982; WATT *et al.* 1983, 1985) provide the best approach for gaining information about current patterns of selection. However, insights into the history of an enzyme polymorphism (as opposed to its current maintenance) must come from analyses of DNA sequence variation underlying protein polymorphism. The two approaches, mechanistic and population genetic/genealogical, are clearly complementary (WATT 1994).

The first study to examine DNA sequence variation underlying protein polymorphism was a survey of variation at the alcohol dehydrogenase (*Adh*) locus in *Drosophila melanogaster* (KREITMAN 1983). Two electrophoretic classes, "slow" and "fast," in *D. melanogaster* show clinal variation in North America (reviewed in CHAMBERS 1988), suggesting that balancing selection is operating on this locus. An analysis of the DNA sequence variation in *Adh* in *D. melanogaster* revealed an excess of polymorphism surrounding the amino acid substitution that distinguishes the slow and fast alleles (HUDSON *et al.* 1987; HUDSON 1990). These data are consistent with a history of balancing selection on the replacement site substitution (HUDSON *et al.* 1987).

However, the pattern of variation for *Adh* in *D. melanogaster* has proved to be exceptional; no evidence of old balanced polymorphism has been found for other polymorphic allozyme loci characterized to date. The slow and fast electrophoretic classes of esterase-6, for exam-

Corresponding author: Laura A. Katz, Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA 01063. E-mail: lkatz@science.smith.edu

ple, show parallel latitudinal clines in *D. melanogaster* and *D. simulans* (ANDERSON and OAKESHOTT 1984), suggesting that balancing selection is operating on shared amino acid polymorphism(s). However, the amino acid replacement substitutions that distinguish the two electrophoretic classes differ between the two species (KAROTAM *et al.* 1995). Furthermore, there is no evidence of inflated levels of polymorphism in esterase-6 in either species, indicating that the parallel clines in the electrophoretic classes are not due to an old balanced polymorphism at a single amino acid replacement site (KAROTAM *et al.* 1995). Likewise, studies of other polymorphic allozyme loci in *Drosophila* such as esterase 5B (VEUILLE and KING 1995), xanthine dehydrogenase (RILEY *et al.* 1992), superoxide dismutase (HUDSON *et al.* 1994) and glucose-6-phosphate dehydrogenase (EANES *et al.* 1993) have failed to show evidence of ancient balancing selection. Either these loci are not evolving under balancing selection, the selection has been too recent to leave a footprint in the DNA, or our current tests for detecting balancing selection are inadequate.

Numerous electrophoretic surveys of variation in the glycolytic enzyme phosphoglucose isomerase (PGI; E.C. 5.3.1.9; also known as glucose-6-phosphate isomerase) have implicated PGI as being under some form of balancing selection (reviewed in RIDDOCH 1993). In several groups of organisms, striking patterns of variation have provided the motivation for examining kinetic differences among PGI genotypes. Most notable are the studies of variation in PGI in *Colias* butterflies, in which WATT and colleagues have elegantly demonstrated that genotypic differences have clear fitness consequences in the field (see WATT 1994). Studies of a variety of other organisms have shown that PGI genotypes differ in kinetic parameters (HOFFMAN 1981, 1983; HALL 1985; ZERA 1987) and in their fitness (performance) in field and laboratory studies (*e.g.*, CARVALHO 1988; PATARNELLO *et al.* 1989; JOHANNESSON *et al.* 1990). Evidence for selection on PGI has also been detected in an analysis of the nucleotide variation underlying electrophoretic classes of PGI from a species of wild yam, *Disocorea tokoro* (TERAUCHI *et al.* 1995). In this study, an excess of nonsynonymous substitutions to synonymous substitutions indicates that the protein is not evolving neutrally.

Here we examine sequence polymorphism in the *Pgi* gene of two sympatric cricket species in which patterns of parallel allelic variation suggest the action of ancient balancing selection maintaining a trans-specific polymorphism. The field crickets, *Gryllus veletis* and *G. pennsylvanicus*, are morphologically similar, broadly sympatric and share similar calling songs (ALEXANDER and BIGELOW 1960). However, they differ in life cycle: *G. veletis* overwinters as a nymph and reproduces in the spring and *G. pennsylvanicus* overwinters as an egg and reproduces in the late summer or early fall. Because of

their similarities, *G. veletis* and *G. pennsylvanicus* were originally thought to be sister species (ALEXANDER and BIGELOW 1960). However, data from surveys of both allozyme variation (HARRISON 1979) and mitochondrial restriction fragment length polymorphisms (RFLPs) (HARRISON and BOGDANOWICZ 1995) indicate that these two species are not close relatives within the North American field crickets.

Surprisingly, *G. veletis* and *G. pennsylvanicus* share all six electrophoretic mobility classes of PGI, whereas at other polymorphic allozyme loci, the two species are either fixed for different alleles or share only one or two alleles (HARRISON 1979; KATZ 1996). Moreover, in a survey of crickets from Virginia to New York, allele frequencies covaried in sympatric populations of the two species for both the PGI-100 and PGI-65 alleles; allele frequencies in populations from a given site are more similar than would be expected by chance (HARRISON 1977a; this study). Both the observation of shared alleles and the covariation in frequencies of the two most common alleles suggest that some form of balancing selection is maintaining variation in PGI in both cricket species.

To gain insights into the origin and history of the shared protein polymorphisms, we sequenced multiple alleles from both *G. veletis* and *G. pennsylvanicus*. The crickets we sampled were all homozygotes, representing four of the shared electrophoretic mobility classes. Although we are missing some sequence from the 5' end of the gene in most crickets, amino acid substitutions consistent with the observed differences in mobility have been identified in both species for all but one electrophoretic class. The substitutions that distinguish the less common alleles from the PGI-100 allele are not the same in *G. veletis* and *G. pennsylvanicus*. Genealogical and population genetic analyses of *Pgi* nucleotide variation suggest that PGI is not evolving neutrally in these two species.

MATERIALS AND METHODS

Crickets of both species were collected from sites in New York, Connecticut and Vermont from 1992 to 1995 (Table 1). To characterize PGI electrophoretic mobility classes, a single hind leg was removed from each cricket and homogenized in water. The homogenates were run on cellulose acetate gels and stained for PGI (HEBERT and BEATON 1993). The PGI mobility classes (PGI-25, PGI-65, PGI-100, PGI-125, PGI-135 and PGI-160) are numbered according to their relative mobility as described in HARRISON (1977b). Individuals homozygous for each mobility class were either collected in the field or generated by crossing field-collected heterozygotes and scoring the progeny.

Table 2 lists the names and collection sites of the 29 crickets (15 *G. veletis* and 14 *G. pennsylvanicus*) characterized at the DNA level in this study. The crickets are named by species (VEL for *G. veletis* and PEN for *G. pennsylvanicus*) and by electrophoretic class. PEN6-100, for example, is the sixth *G. pennsylvanicus* from which RNA was extracted and is homozygous for the PGI-100 allele.

TABLE 1
Collection sites for crickets of both species

Site	State	County	Description
HC-Moun	CT	Hartford	Mountain Rd., Farmington
LC-Hous	CT	Litchfield	Rte. 5, Cornwall Bridge
LC-Thom	CT	Litchfield	Black Rock dam, Thomaston Bridge
NHC-Derby	CT	New Haven	Rte. 7, Derby
BC-Snow	NY	Broome	Rte. 79, Colesville
OC-17K	NY	Orange	Rtes. 17K and 84
PC-Kent	NY	Putnam	Rte. 301, Kent Cliffs
TC-Besemer	NY	Tompkins	Besemer Hill Rd., Ithaca
TC-Fire	NY	Tompkins	Fire Tower Rd., Dryden
TC-Hung	NY	Tompkins	Hungerford Rd., Ithaca
TC-MtPl	NY	Tompkins	Mt. Pleasant Rd., Farmington
TC-Turk	NY	Tompkins	Turkey Hill Rd., Ithaca
WaC-Whit	NY	Washington	Rte. 45, Whitehall
WC-NorD	VT	Windsor	The dump, Norwich
WC-OldF	VT	Windsor	Old Farm Rd., Norwich
WC-Rte5	VT	Windsor	Rte. 5, Norwich
WC-Tabo	VT	Windsor	Tabor Rd., Norwich
WC-Tree	VT	Windsor	New Boston Rd., Norwich

Each site is identified first by county and then by an abbreviation of the road or landmark near the site. Further descriptions of sites can be found in KATZ (1996).

The entire coding region of the *Pgi* gene has been described for both species of crickets (KATZ 1997). Because of the presence of large introns, RT-PCR was used to characterize RNA molecules from each species. Total RNA was prepared following CHOMCZYNSKI and SACCHI (1987), and first strand cDNAs were generated by reverse transcription using either an oligo-dT primer or a gene-specific primer. These cDNAs were subsequently amplified and sequenced by end-labeling primers with ^{32}P and cycle sequencing.

Despite repeated attempts at both the reverse-transcription and amplification steps, complete sequence could not be obtained from the 5' end of the *Pgi* gene. Failure to consistently amplify a portion of the 5' end of the gene suggests that either there is substantial secondary structure in the RNA

molecules in this region or a portion of this region is alternatively spliced *in vivo*. Attempts to melt secondary structure by performing reverse transcriptions at 55°, the upper range of temperature at which reverse transcriptase will work, allowed successful characterization of the 3' end of the gene but did not affect the ability to amplify the 5' end. The missing data from each individual are indicated in Table 2.

Sequences for population analysis were generated directly from PCR amplified fragments using template DNA from individuals homozygous for single electrophoretic mobility classes. These PCR products were sequenced using end-labeled primers and protocols adapted from Stratagene, Inc. as described elsewhere (KATZ 1997). Because homozygosity of the crickets was defined by electrophoretic mobility, allelic

TABLE 2
Crickets sequenced in this analysis

Cricket	Location	Missing amino acids	Cricket	Location	Missing amino acids
VEL3-100	OC-17k	47-139	PEN3-100	TC-Fire	79-162
VEL4-65	OC-17k	1-150	PEN6-100	WC-Tre	1-162
VEL5-100	NHC-Derb	55-159	PEN7-100	WC-Tabo	82-107
VEL6-65	NHC-Derb	1-162	PEN9-65	TC-Fire	82-107
VEL7-65	OC-17k		PEN10-135	WC-Tabo	82-107
VEL8-65	NHC-Derb	66-115	PEN11-100	WC-Tabo	1-87, 124-159
VEL9-100	NHC-Derb	82-107	PEN12-25	WC-Tabo	1-55, 124-157
VEL10-100	TC-Bese	82-107	PEN13-135	TC-Hung	1-55, 124-158
VEL12-100	NHC-Derb	66-67, 75-157	PEN14-65	WC-Tabo	1-55, 124-160
VEL13-125	NHC-Derb	1-157	PEN17-100	WC-Tabo	1-155
VEL14-135	NHC-Derb	39-157	PEN18-100	LC-Hous	1-155
VEL16-65	OC-17k	1-297	PEN19-135	TC-Genu	1-155
VEL17-100	NHC-Derb	1-151	PEN20-135	TC-Fire	1-155
VEL18-100	WaC-Whit	1-151	PEN21-25	WC-Tabo	1-155
VEL21-125	PC-Kent	1-151			

Each cricket is identified by species, a number referring to the order in which the cDNAs were prepared and by electrophoretic genotype. Location refers to collection site (Table 1). The last column indicates amino acids not characterized for individual (see MATERIALS AND METHODS).

TABLE 3
PGI allele frequencies in *G. veletis* and *G. pennsylvanicus* from 1974 to 1976 and 1992 to 1995

Allele	<i>G. vel</i>		<i>G. pen</i>	
	1992–1994	1974–1976	1992–1994	1974–1976
25	0.0	0.8	3.5	0.0
65	9.6	11.3	7.2	8.1
100	79.0	80.0	74.2	79.5
125	5.9	4.5	1.8	1.1
135	4.8	3.3	13.0	11.8
160	0.4	0.2	0.2	0.0
No. of crickets	136	230	299	247

Frequency of PGI alleles as determined by protein gel electrophoresis. Crickets from 1992 to 1994 (this study) were collected in New York, Connecticut, and Vermont. Crickets from 1974 to 1976 were collected in New York, Connecticut, and Virginia (HARRISON 1979).

classes may exhibit heterozygosity in the underlying DNA sequences. For the purpose of the analysis, any nucleotide position that appeared heterozygous in only one individual on only one sequencing gel is considered to be due to *Taq* polymerase error and is excluded from the data set. This exclusion is conservative with respect to the hypothesis that balancing selection is operating on this locus in that it will result in an underestimate of levels of polymorphism in the two species.

Sequences were entered and aligned by eye using LaserGene software (DNASTAR 1994). Phylogenetic analyses were performed using maximum parsimony and neighbor-joining algorithms (PAUP test version 4.0D55: SWOFFORD 1995). HKA tests (HUDSON *et al.* 1987) were done using a program written by M. FORD (Northwest Fisheries Science Center, Seattle, WA) and B. NOON (Cornell University, Ithaca, NY).

RESULTS

Allozyme variation: Electrophoretic variation in PGI was determined for 299 *G. pennsylvanicus* and 136 *G. veletis* individuals collected between 1992 and 1994 (KATZ 1996). Six electrophoretic classes were detected in each species, five of which were shared between the species (Table 3). A unique allele, named PGI-75, was detected in one *G. veletis* individual. The frequencies of the shared alleles in both species are very similar to each other and to the average frequencies found in the original allozyme survey done more than 20 years ago (Table 3). In fact, although only four sample localities are common to both studies, the frequencies of the four most common alleles within each species (PGI-100, PGI-65, PGI-135, PGI-125) are not statistically different ($\chi^2 = 2.12$, $P = 0.547$ for *G. veletis* and $\chi^2 = 2.45$, $P = 0.484$ for *G. pennsylvanicus*). Averaged across all populations, the frequencies of the three most common alleles are in Hardy-Weinberg equilibrium in both species ($\chi^2 = 1.18$, $P = 0.947$ for *G. veletis* and $\chi^2 = 5.23$, $P = 0.388$ for *G. pennsylvanicus*). A Ewens-Watterson test (WATTERSON 1975) detects no evidence of a nonneutral distribution of alleles within either species.

Based on data from 1974 to 1976, HARRISON (1977a) showed that, for both the PGI-100 and PGI-65 alleles, there was a significant rank correlation between allele

frequencies in the two species when sympatric populations were sampled from North Carolina, Virginia, New York and Connecticut. In contrast to the 1974–1976 data, there is no significant correlation between species of the transformed frequencies of either the PGI-100 or PGI-65 alleles in the four sympatric populations surveyed for this study (data not shown). However, the lack of a correlation is not surprising given the small number of individuals collected at each site ($n = 4–37$) and the small number of populations being compared.

To further examine the pattern of covariation in both the PGI-100 and PGI-65 electrophoretic classes detected from crickets collected in 1974–1976 (HARRISON 1977a), we have reanalyzed the original allozyme data (Table 4) using a log linear model. The advantage of this model over a regression analysis is that it (1) allows us to eliminate higher order interactions as explanations for the patterns in the data; (2) does not make any assumptions about the variances for the populations sampled; and (3) is more appropriate when sample sizes are not equal. We built a model in SYSTAT (1996) from observed numbers of four relatively common and widespread alleles (PGI-65, PGI-100, PGI-125, and PGI-135). This model detects a marginally significant three-way interaction of allele $\times$ species $\times$ location ($P = 0.058$), indicating that further consideration of significant deviations in the data are inappropriate. (This is true whether we treat the 125 and 135 alleles separately or if we combine these alleles into one allelic class.) However, an analysis of the PGI-65 and PGI-100 alleles alone indicates that there is no significant three-way interaction ($P = 0.29$). More importantly, the allele $\times$ species interaction does not contribute significantly to the model as the removal of this term does not significantly affect the overall χ^2 (change in $\chi^2 = 0.95$, $P = 0.33$). This implies that there is no significant difference between species in the frequencies of these alleles in sympatric populations. In contrast, the species $\times$ location and allele $\times$ location interactions are significant, indicating that there is significant variability between spe-

TABLE 4

Species	Allele	Shar	VA	BD1	BD2	BD4	BD5	BR2	BR4	BR5	BR6	BR7
<i>Gve</i>	65:100	2:51 (0.039)	5:27 (0.185)	12:61 (0.197)	9:68 (0.132)	7:62 (0.113)	10:75 (0.133)	3:28 (0.107)	10:91 (0.109)	7:83 (0.084)	15:104 (0.144)	6:53 (0.113)
<i>Gpe</i>	65:100	1:109 (0.009)	4:26 (0.154)	23:130 (0.177)	13:74 (0.176)	17:122 (0.139)	10:106 (0.094)	2:59 (0.034)	3:49 (0.061)	0:65 (0.000)	8:78 (0.103)	14:63 (0.222)

Gve, *G. veletis*; *Gpe*, *G. pennsylvanicus*; numbers in parentheses are the ratio of 65:100. Populations are from Litchfield Co., Connecticut (Shar); Tompkins Co., New York (VA, BD1, BD2, BD4, BD5); Rockbridge Co., Virginia (BR2, BR4, BR5); and Alleghany Co., North Carolina (BR6, BR7). Further description of population locations are in HARRISON (1977b).

cies and location (change in $\chi^2 = 96.97$ and 40.69, respectively; $P < 0.001$ for both). Hence, the log-linear models indicate that the frequencies of the PGI-65 and PGI-100 alleles do not differ between sympatric popula-

tions of the two species although frequencies of these alleles clearly vary geographically.

Variable sites in the cDNA: A total of 37 variable nucleotide sites was identified from sequence analysis

TABLE 5

	-	1	1	2	2	3	6	6	7	7	7	8	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
--	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

-, missing data; ., matches consensus *G. pennsylvanicus* sequence; S, synonymous substitution; R, replacement substitution; *, charge-changing replacement substitution. Phylogenetic and population genetic analyses use nucleotides 486 through the 3' untranslated region. Polymorphic DNA designations are indicated using standard IUB-IUPAC code.

FIGURE 1.—Distribution of polymorphic sites. The horizontal line represents the gene, with the thicker line being the region of virtually complete data. Numbers above the line are for *G. veletis* and below the line are for *G. pennsylvanicus*. S, synonymous substitution; R, replacement substitution; *, charge-changing substitution. Gray lines indicate the two shared polymorphisms.

of *Pgi* from 29 crickets (Table 5). Each variable site along the cDNA is numbered with the initiation codon assigned a position of 1 (Figure 1 and Table 5). The full intron/exon structure of *Pgi* in *G. veletis* and *G. pennsylvanicus* has not been determined. However, genomic DNA, characterized from the region between nucleotides 966 and 1110 in the cDNA, reveals that there are two large introns totaling ~1750 bp (KATZ 1997). Consequently, sites reported in Table 5 may be more distant from one another along the chromosome than their positions in the cDNA sequence suggest.

One of the variable sites (-31) is in the 5' untranslated region and three (+1, +2, and +25) are in the 3' untranslated region (Figure 1). Many of the observed variants are rare, appearing in only one or two sequences. Only two fixed differences were found between the two species, at positions -31 and 119. Both of these sites are found in the 5' region where data are missing from most individuals.

Amino acid substitutions and electrophoretic mobility: Although data are missing from the 5' end of the gene (Table 2), amino acid replacements consistent with differences in mobility were identified for all but the *G. pennsylvanicus* PGI-135 electrophoretic class (Table 6). Shown in bold in Table 6 are charge-changing amino acid replacements, which appear to be responsible for observed changes in electrophoretic mobility of the less common PGI alleles relative to the PGI-100 alleles. These changes are based on identification of replacement substitutions to or from one of the four charged amino acids [arginine (R) = +1; lysine (K) = +1; aspartic acid (D) = -1; and glutamic acid (E) = -1].

In *G. veletis*, a single charge-changing amino acid substitution was found for each electrophoretic class. For example, the single PGI-135 allele sampled (VEL14-135) has a threonine (T) replacing a lysine (K) at amino acid position 242 (Table 6). This substitution gives the PGI-135 allele a net charge of -1 relative to the 100 alleles, consistent with the faster anodal migration of this allele relative to PGI-100 on cellulose acetate gels. Both of the *G. veletis* PGI-125 alleles characterized have a glutamine (Q) substituted for an arginine (R) at posi-

tion 206, giving these alleles a net charge of -1. Either conformational changes associated with the charge-changing substitutions in the *G. veletis* PGI-125 and PGI-135 alleles or undetected amino acid substitutions in the 5' end of the gene must be invoked to explain the small difference in mobility of these two alleles. For the *G. veletis* PGI-65 alleles, the VEL7-65 individual has an arginine (R) substituted for a glutamine (Q), consistent with the slower migration of this allele. The second *G. veletis* individual for which data are available in this region (VEL8-65) was ambiguous at this site on the sequencing gel.

In *G. pennsylvanicus* the pattern is more complicated. A single amino acid replacement appears to define the PGI-25 mobility class (Table 6). However, the PGI-65 class is defined by at least two different charge-changing amino acid substitutions: a lysine (K) substituted for a glutamine (Q) at amino acid position 543 in PEN14-65 or an asparagine (N) for an aspartic acid (D) at amino acid position 71 in PEN9-65. Both of these substitutions result in a net charge change of +1, consistent with the slower mobility of PGI-65 relative to PGI-100. No charge changing substitutions were observed in the PGI-135 class of *G. pennsylvanicus* although amino acid substitutions presumably occur in the 5' end of these alleles for which no sequence is available (see Table 2).

Pairwise differences between sequences: Phylogenetic and population genetic analyses of *Pgi* sequence variation were carried out using virtually complete sequences downstream of amino acid 162 plus 30 bp of the 3' untranslated region. Across this region, the average pairwise difference ($d_{x,y}$) between alleles within species is higher for *G. pennsylvanicus* (3.44) than for *G. veletis* (1.31). Comparisons within and between electrophoretic classes indicate that most of the variation in the *G. veletis* sample is due to differences between the rare and common alleles (Table 7). For *G. veletis*, $d_{x,y}$ is ≤ 1 for comparisons within electrophoretic mobility classes and between the PGI-65 and PGI-100 classes. In contrast, $d_{x,y}$ is always greater than two for comparisons between either the PGI-125 or the PGI-135 alleles and all other classes. This pattern does not hold for *G. pennsylvanicus* for which the largest value for $d_{x,y}$ (12.0) rep-

TABLE 6
Amino acid replacements by charge class

Allele	Cricket	Amino acid position ^a									
		31 Q	40 N/S	71 D	206 R	242 K	282 Y	302 E	455 A	468 A	543 Q
100	PEN3	.	N	.	.	.	.	.	.	.	.
100	PEN6	?	?	?	.	.	.	.	.	.	.
100	PEN7	.	N	.	.	.	.	.	.	.	.
100	PEN11	?	?	?	.	.	.	.	.	.	.
100	PEN17	?	?	?	.	.	.	.	.	.	.
100	PEN18	?	?	?	.	.	.	.	.	.	.
100	VEL3	.	S	?	.	.	.	.	A/S	.	.
100	VEL5	.	S	?	.	.	.	.	A/S	.	.
100	VEL9	.	S	.	.	.	.	.	A/S	.	.
100	VEL10	.	S	.	.	.	.	.	.	.	.
100	VEL12	.	S	.	.	.	.	.	S	.	.
100	VEL17	?	?	?	.	.	.	.	.	.	.
100	VEL18	?	?	?	.	.	.	.	.	.	.
125	VEL13	?	?	?	Q	.	.	.	S	.	.
125	VEL21	?	?	?	Q	.	Y/H	.	S	.	.
135	PEN10	.	N	.	.	.	.	.	.	.	.
135	PEN13	?	?	.	.	.	.	.	.	.	.
135	PEN19	?	?	?	.	.	.	.	.	.	.
135	PEN20	?	?	?	.	.	.	.	.	T	.
135	VEL14	.	S	?	.	T	.	.	.	.	.
25	PEN12	?	?	.	.	.	.	K	.	.	.
25	PEN21	?	?	?	.	.	.	K	.	.	.
65	PEN9	.	N	N	.	.	.	.	.	.	.
65	PEN14	?	?	.	.	.	.	.	.	.	K
65	VEL4	?	?	?	.	.	.	.	.	.	.
65	VEL6	?	?	?	.	.	.	.	A/S	.	.
65	VEL7	R	S	.	.	.	.	.	A/S	.	.
65	VEL8	Q/R	S	?	.	.	.	.	A/S	.	.
65	VEL16	?	?	?	?	?	.	.	.	.	.

Charge changing amino acid substitutions are relative to the PGI-100 alleles (and the consensus) and are defined by the gain or loss of a charged amino acid [arginine (R), lysine (K), glutamic acid (E) and aspartic acid (D)]. Bold letters are charge changing substitutions. ?, missing data; /, polymorphic or ambiguous.

^aConsensus appears below amino acid position number.

resents the difference between two PGI-65 alleles (Table 7), consistent with the independent origins of these alleles (Table 6). Moreover, four *G. pennsylvanicus* alleles, which have the highest values of d_{xy} when compared with all other alleles, represent three electrophoretic classes: PEN9-65 ($d_{xy} = 7.38$, compared to all other alleles), PEN14-65 ($d_{xy} = 5.62$), PEN6-100 ($d_{xy} = 5.62$), PEN20-135 ($d_{xy} = 4.62$).

Genealogical analyses: A strict consensus tree generated from a heuristic search using the maximum parsimony algorithm in PAUP (test version 4.0.D55; SWOFFORD 1995) shows very little resolution (Figure 2a). In an unrooted genealogy, most of the *G. pennsylvanicus* sequences form a single exclusive group. An exclusive group is "one whose members are all more closely related to each other than to any ... [members] outside the group" (BAUM and SHAW 1995).] However, the

PEN9-65 haplotype appears more closely related to a *G. veletis* sequence than to the other *G. pennsylvanicus* sequences. Also, although the relationships are unresolved, two other *G. pennsylvanicus* alleles (PEN6-100 and PEN10-135) do not appear to fall within the *G. pennsylvanicus* clade. In the parsimony analysis, electrophoretic classes do not form exclusive groups within either species with the exception of the two PGI-25 alleles in *G. pennsylvanicus* and the two PGI-125 alleles in *G. veletis*.

A genealogy constructed using neighbor-joining (Figure 2b; PAUP test version 4.0.D55; SWOFFORD 1995) is consistent with the observations from the maximum parsimony tree (Figure 2a). However, in the neighbor-joining tree, both PEN9-65 and PEN6-100 clearly fall within the *G. veletis* clade (Bootstrap value = 95%). Also, in the distance-based analysis an additional elec-

TABLE 7
Average pairwise differences between alleles
for both species

	PGI-65	PGI-100	PGI-125	PGI-135
<i>G. veletis</i>				
PGI-65	0.80 (10)	0.51 (35)	2.90 (10)	2.80 (5)
PGI-100		0.33 (21)	2.56 (14)	2.71 (7)
PGI-125			1.00 (1)	4.50 (2)
PGI-135				—
<i>G. pennsylvanicus</i>				
PGI-65	12.0 (1)	5.50 (12)	6.38 (8)	7.00 (4)
PGI-100		2.33 (15)	2.54 (24)	2.50 (12)
PGI-135			2.67 (6)	2.50 (8)
PGI-25				0.00 (1)

Calculated across the region for which we have virtually complete data (nucleotide 486 through the 3' untranslated region). Numbers along the diagonal are comparisons within electrophoretic classes. Parentheses contain the number of comparisons made. Calculated excluding polymorphic sites.

trophoretic class, the PGI-135 alleles in *G. pennsylvanicus*, forms an exclusive group. Strikingly, neither the PGI-100 nor the PGI-65 alleles form exclusive groups within either species using neighbor-joining or maximum parsimony.

Departures from neutrality: An HKA test (HUDSON *et al.* 1987) was used to determine if the pattern of variation in *Pgi* within and between *G. pennsylvanicus* and *G. veletis* is significantly different from that observed for a second region of the cricket genome. This test is based on the prediction that under neutrality, the ratio of polymorphism within a species to divergence between species should be the same for all loci. Specifically, the test compares the ratio of the number of segregating sites within a species (S) to the average pairwise differences between alleles of the two species ($d_{v,p}$). The assumptions of the HKA test are quite rigid, requiring a random sample of alleles from populations that are at equilibrium, where generations are distinct, mutations are distributed in a Poisson fashion, and recombination does not occur (for a further discussion of assumptions, see HUDSON *et al.* 1987). As discussed below, some of these assumptions are likely to be violated in the current study making the results of the HKA test suggestive but not conclusive.

No other nuclear genes have been characterized from these field crickets so it is only possible to compare *Pgi* to mitochondrial genes. A total of 1621 bp representing a portion of the cytochrome oxidase I gene, a tRNA, and a portion of cytochrome oxidase II have been sequenced from six northern *G. pennsylvanicus* and one *G. veletis* (C. S. WILLETT, M. FORD and R. G. HARRISON, unpublished results). Because polymorphism data are only available for mtDNA from *G. pennsylvanicus*, polymorphism in *Pgi* from *G. veletis* was not included in this analysis (Table 8). An HKA test compar-

FIGURE 2.—*Pgi* gene genealogies. (a) Strict consensus of 1356 most parsimonious trees of length 24 constructed using 10 random addition sequences in a heuristic search (PAUP, test version 4.0D55; SWOFFORD 1995) using nucleotides 542 through the beginning of the 3' untranslated region. Multistate positions are treated as uncertainties. (b) A neighbor-joining tree constructed in PAUP (test version 4.0D55; SWOFFORD 1995) using uncorrected nucleotide distances. Ambiguous sites are ignored. Numbers represent bootstrap values from 100 replicates. Branches without numbers have bootstrap values <50%. Both trees are unrooted.

ing the ratio of the number of segregating sites (S) to the average pairwise differences between the species ($d_{v,p}$) was significant ($\chi^2 = 6.85$, $P = 0.012$), indicating that at least one of these genes is not evolving neutrally.

Another prediction of the neutral theory is that under mutation-drift equilibrium, the number of segregating sites (S) and the average pairwise difference between alleles in a sample ($d_{x,y}$) can both be related to $4N_d\mu$ in a predictable way (Tajima 1989). Tajima's D statistic, which tests the differences in estimates of $4N_d\mu$ from observed values of (S) and ($d_{x,y}$), is negative for both species (-1.42 for *G. veletis* and -1.29 for *G. pennsylvanicus*). A negative D indicates an excess of segregating sites relative to the average pairwise difference between alleles, but neither of these estimated values is significant.

TABLE 8
HKA test of ratio of polymorphism to divergence for *Pgi* and mitochondrial DNA

	S-PEN	$d_{v,p}$	No. of base pairs	χ^2	Significance
<i>Pgi</i>	16 (14)	6.03	1229		
mtDNA ^a	5 (7)	85.43	1621	6.85	$P = 0.012$

S, number of segregating sites by species; $d_{v,p}$, average pairwise differences between species. Parentheses contain number of *G. pennsylvanicus*.

^a Mitochondrial sequence data from C. S. WILLETT, M. FORD and R. G. HARRISON (unpublished results).

DISCUSSION

Molecular basis of allozyme variants: Possible explanations for the sharing of electrophoretic mobility classes and observed covariation in the frequency of both the PGI-100 and PGI-65 electrophoretic classes in sympatric populations (HARRISON 1977a; log linear analysis, this study) are as follows: balancing selection operating in both species in parallel; recent sympatric speciation; and current introgression. Recent divergence and current introgression predict that alleles representing the same electrophoretic mobility class from the two species should have a very recent common ancestor; they should be more closely related to each other than to different alleles within the same species. Similarly, most models of trans-specific balancing selection make the prediction that shared electrophoretic classes should share the same replacement site substitutions. However, we find that identical electrophoretic mobility classes in the two species are not diagnosable by the same amino acid substitutions (Table 6) and that electrophoretic classes do not form exclusive groups in phylogenetic analyses (Figure 2). Thus, arrays of alleles with the same mobility have apparently arisen independently in the two species (and at least in the case of PGI-65 in *G. pennsylvanicus*, have arisen more than once within a species). Consequently, sharing of alleles due to recent divergence or introgression can be excluded as explanations for patterns of variation for PGI. If balancing selection is acting to maintain variation, then it must be acting on the charge state of the enzyme or some other character associated with the electrophoretic mobility of the protein molecule. Why this might be the case is not clear, but a role for balancing selection is supported by analysis of the patterns of nucleotide variation and of the ratios of conservative to radical amino acid substitutions in the two species (see below).

In a survey of patterns of variation in PGI from more than 75 taxa, RIDDOCH (1993) detected a significant correlation between anodal forms of PGI and what he presumed to be "stressful" environments, such as increased temperature or increased salinity. On this basis, he argued that a trade-off between stability of the protein and catalytic efficiency is associated with charge class and that such a trade-off could maintain variation in the protein (RIDDOCH 1993). Our data are consistent with this hypothesis, but information on the biochemi-

cal properties of the allozymes would be needed to demonstrate that molecules with similar net charge (electrophoretic mobility) have similar properties despite differences in amino acid substitutions responsible for mobility differences.

Non-charge-altering amino acid substitutions that change the conformation of the protein could also lead to a change in the electrophoretic mobility of PGI. However, no such amino acid substitutions diagnose electrophoretic classes (Table 6). It is also possible that nucleotide sequence from the 5' end of the gene will reveal additional charge-changing amino acid substitutions. But, such substitutions would presumably require compensating replacement substitutions if variation in net charge is to remain consistent with electrophoretic mobility differences.

Evidence of selection from nucleotide variation: Patterns of variation in the cDNA sequences within and between species are consistent with the hypothesis that PGI is under some form of balancing selection. Most of the variation in the cricket *Pgi* gene partitions within species (Table 9, Figure 2b). Within the coding region of *G. veletis* there are 11 polymorphic sites (five replacement and six synonymous) while within *G. pennsylvanicus* there are 20 polymorphic sites (five replacement substitutions and 15 synonymous substitutions). Surprisingly, there are few fixed differences between the cricket species and few shared polymorphisms. There is one fixed replacement site substitution at position 119 and two shared polymorphic synonymous substitutions at positions 657 and 1506 (Table 5). Variation appears to have accumulated independently in the two species without the parallel accumulation of divergence. This pattern is consistent with balancing selection acting on *Pgi* electrophoretic mobility classes in subdivided populations.

The inflated level of polymorphism within species is due, in part, to the presence of *G. veletis*-like alleles in *G. pennsylvanicus* (Figure 2). Two *G. pennsylvanicus* alleles, PEN9-65 and PEN6-100, share five synonymous substitutions (positions 1122, 1461, 1518, 1587 and 1656) with *G. veletis* (Figure 2 and Table 5). While the pattern of variation might reflect recent introgression between species, available data suggest that this is an unlikely explanation. First, *G. veletis* and *G. pennsylvanicus* are fixed for different electrophoretic alleles of pho-

TABLE 9

Comparison of patterns of radical and conservative amino acid substitutions within and between crickets and *Drosophila*

Substitutions	Polymorphic within <i>G. veletis</i>	Polymorphic within <i>G. pennsylvanicus</i>	Shared between species	Fixed between species
Crickets				
Replacement	5	5	0	1
Conservative	1	2	0	1
Radical	4	3	0	0
Synonymous	6	15	2	0
Substitutions	Polymorphic within <i>D. yakuba</i>	Polymorphic within <i>D. melanogaster</i>	Polymorphic within <i>D. simulans</i>	Fixed between species
Drosophila				
Replacement	1	2	2	11
Conservative	1	2	2	7
Radical	0	0	0	4
Synonymous	12	2	19	85

Drosophila data from J. H. McDONALD (unpublished data).

sphoglucomutase, and no heterozygotes have been observed in extensive samples from a variety of localities where the two species are sympatric (HARRISON 1979). Second, attempts to hybridize these two species in the laboratory have consistently failed (BIGELOW 1958, 1960; ALEXANDER and BIGELOW 1960).

Alternatively, the presence of *G. veletis*-like sequences in *G. pennsylvanicus* could result from persistence of an ancestral polymorphism. Ancestral polymorphism can persist if there has been insufficient time for lineage sorting to occur or balancing selection maintains alleles within a population. Both distance and phylogenetic analysis of allozyme and mtDNA RFLP data indicate that the two cricket species are not closely related within the genus *Gryllus*. In surveys of 16 allozyme loci, comparisons among four populations of *G. pennsylvanicus* and six populations of *G. veletis* gave values of Nei's *D* of 0.002–0.022 and 0.003–0.011, respectively, while Nei's *D* between the two species is 0.160 (HARRISON 1979). Similarly, an RFLP survey of mtDNA from four individuals of each species indicated that there is 0% divergence among *G. pennsylvanicus* sequences and 1.5% divergence among *G. veletis* sequences, while the divergence between the two species was estimated to be 6% (HARRISON and BOGDANOWICZ 1995). Given that the divergence between *G. veletis* and *G. pennsylvanicus* must be relatively old, persistence of neutral ancestral polymorphism in *Pgi* is considered unlikely.

A HKA test of the ratio of divergence to polymorphism in mtDNA and *Pgi* sequences in *G. pennsylvanicus* shows a significant departure from expectations under neutrality ($P = 0.012$; Table 8). There are two likely explanations for this departure: increased polymorphism in *Pgi* or reduced polymorphism in cricket mtDNA. Patterns of variation from the original allozyme survey (HARRISON 1979) are consistent with the pattern observed for the mitochondrial sequence data

and not for *Pgi*. For both mtDNA and allozymes, *G. veletis* and *G. pennsylvanicus* appear to form exclusive groups with considerably more variation between the two species than within either species. This suggests that the HKA test is significant because of inflated levels of *Pgi* polymorphism within species relative to divergence between species. Such a conclusion is consistent with expectations if balancing selection is operating on the *Pgi* locus.

Caution must be taken in interpreting the HKA test because of the nonrandom manner in which the alleles were sampled and because of uncertainties about the population structure of these crickets. Sampling may not be a problem for the *G. pennsylvanicus* sequences since variation among alleles does not partition by electrophoretic classes (see RESULTS and Table 7). Perhaps more worrisome is the assumption in both the HKA and TAJIMA's tests that the populations are at equilibrium (HUDSON *et al.* 1987; TAJIMA 1989). While the level of population subdivision is not known for *G. pennsylvanicus* and *G. veletis*, the crickets live in disturbed habitats and probably experience frequent extinctions and recolonizations. As a result, populations of these two species are not likely to be in equilibrium. Finally, if balancing selection has been operating in the two species since before their divergence, the HKA test may no longer be appropriate. The HKA test is designed to detect balancing selection as an increase in the ratio of polymorphism to divergence in one locus when compared with a neutrally evolving locus. However, trans-specific balancing selection will increase both the polymorphism within a species and the level of divergence between species.

Patterns of substitutions: Patterns of conservative and radical amino acid substitutions also suggest that *Pgi* is not evolving neutrally in *G. veletis* and *G. pennsylvanicus* (Tables 9 and 10). Radical substitutions are de-

TABLE 10
Amino acid substitutions in sample

Codon position ^a	DNA	Amino acid	Charge	Size	Hydrophobicity	Substitution	Distribution in sample ^a
92	CAG CGG	Q R	None +	Large Large	Hydrophilic Hydrophilic	Radical	All others One <i>G.v.</i> 65 (7) ambiguous in the other <i>G.v.</i> 65 (8)
119	AAC AGC	N S	None None	Small Tiny	Hydrophilic Hydrophilic	Conservative	All <i>G.p.</i> All <i>G.v.</i>
211	AAT GAT	N D	None –	Small Small	Hydrophilic Hydrophilic	Radical	All others One of two <i>G.p.</i> 65 (9)
617	CGA CAA	Q R	None +	Large Large	Hydrophilic Hydrophilic	Radical	All others Both <i>G.v.</i> 125 (13, 21)
725	AAG ACG	K T	+ None	Large Small	Hydrophobic Hydrophobic	Radical	All others <i>G.v.</i> 135 (14)
736	ACA GCA	T A	None None	Small Tiny	Hydrophobic Very hydrophobic	Conservative	All others One of two <i>G.P.</i> 65 (14), ambiguous in one <i>G.P.</i> 100 (18)
844	TAC CAC	Y H	None (+)	Large Large	Hydrophobic Hydrophobic	Conservative	All others Ambiguous in one of two <i>G.v.</i> 125 (21)
904	GAA AAA	E K	– +	Large Large	Hydrophilic Hydrophobic	Very radical	All others All <i>G.p.</i> 25s (12, 21)
1363	GCA TCA	A S	None None	Tiny Tiny	Very hydrophobic Hydrophilic	Radical	All others Both <i>G.v.</i> 125 (13, 21); One <i>G.v.</i> 100 (12); ambiguous in three <i>G.v.</i> 100s (3, 5, 9) and two <i>G.v.</i> 65 (7, 8)
1402	GCC ACC	A T	None None	Tiny Small	Very hydrophobic Hydrophobic	Conservative	All others One <i>G.p.</i> 135 (20)
1627	CAA AAA	Q K	None +	Large Large	Hydrophilic Hydrophobic	Radical	All others One <i>G.p.</i> 65 (14)

Based on designations of amino acid classes by TAYLOR (1986). A size change between small and tiny is considered conservative.
^a Parentheses contain cricket identification number.

defined as changes in the charge, size or hydrophobicity of an amino acid based on characterizations of TAYLOR (1986). The one substitution between the species (amino acid position 40) is a conservative substitution. However, within *G. veletis* and *G. pennsylvanicus*, four of five and three of five amino acid substitutions are radical substitutions, respectively (Table 9). One of the radical substitutions in *G. pennsylvanicus*, the substitution of a positively charged lysine (K) for negatively charged glutamic acid (E) at position 302 (Table 10), is particularly noteworthy. This substitution, which diagnoses the PGI-25 class in *G. pennsylvanicus*, is both biochemically and phylogenetically a radical substitution: all other complete animal and fungi sequences have a glutamic acid at this position (see KATZ 1997).

In contrast to the pattern in crickets, there are nearly two times as many conservative substitutions as radical substitutions between the *D. melanogaster*, *D. yakuba* and *D. simulans* (seven conservative substitutions to four radical substitutions; Table 9). And, within all three Dro-

sophila species, the small number of amino acid replacements are all conservative (Table 9). Inflated levels of radical substitutions relative to conservative substitutions in the crickets could be due to relaxed constraints or, as suggested above, to balancing selection operating on charge states of the PGI protein.

Conclusions: The phenomenon reported here, heterogeneity at the amino acid level within electrophoretic mobility classes that are under balancing selection, may be widespread. If balancing selection is operating on an emergent property of the protein, then current tests for detecting footprints of selection may be inadequate. For example, two esterase-6 electrophoretic classes, which show parallel clines in two species of *Drosophila*, are distinguished by different amino acid substitutions (KAROTAM *et al.* 1995). It is possible that the parallel clines are subject to balancing selection but that our current models, which depend upon shared amino acid polymorphisms, are not adequate for detecting selection. What is needed

are models for detecting selection that can incorporate the biochemistry and the secondary and tertiary structure of proteins.

The authors thank Drs. MONICA GEBER, CHIP AQUADRO, MIKE FORD and two reviewers for reading this manuscript. Also, thanks to STEVE BOGDANOWICZ for technical help, DOUGLAS DEUTSCHMAN for assistance with the log linear model and MIKE FORD, MICHAEL NACHMAN and ALEXEY KONDRASHOV for helpful discussions. This work was supported by National Science Foundation grant DEB 93-18472.

LITERATURE CITED

- ALEXANDER, R. D., and R. S. BIGELOW, 1960 Allochronic speciation in field crickets and a new species, *Acheta veletis*. *Evolution* **14**: 334–346.
- ANDERSON, P. R., and J. G. OAKESHOTT, 1984 Parallel geographical pattern of allozyme variation in two sibling *Drosophila* species. *Nature* **308**: 729–731.
- BAUM, D. A., and K. L. SHAW, 1995 Genealogical perspectives on the species problem, pp 289–303 in *Experimental and Molecular Approaches to Plant Biosystematics*, edited by P. C. HOCH and A. G. STEVENSON. Missouri Botanical Garden, St. Louis, MO.
- BIGELOW, R. S., 1958 Evolution in the field cricket *Acheta assimilis* Fab. *Can. J. Zool.* **36**: 139–151.
- BIGELOW, R. S., 1960 Interspecific hybrids and speciation in the genus *Acheta* (Orthoptera: Gryllidae). *Can. J. Zool.* **38**: 509–524.
- CARVALHO, G. R. 1988 Differences in the frequency and fecundity of PGI-marked genotypes in a natural population of *Daphnia magna* Straus (Crustacea: Cladocera). *Funct. Ecol.* **2**: 453–462.
- CHAMBERS, G. K., 1988 The *Drosophila* alcohol dehydrogenase gene-enzyme system. *Adv. Genet.* **25**: 39–107.
- CHOMCZYNSKI, P., and N. SACCHI, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156–159.
- DIMICHELE, L., and D. A. POWERS, 1982 Physiological basis for swimming endurance differences between Ldh-B genotypes of *Fundulus heteroclitus*. *Science* **216**: 1014–1016.
- DNASTAR, 1994 Lasergene. DNASTAR Inc., Madison, WI.
- EANES, W. F., M. KIRCHNER and J. YOON, 1993 Evidence for adaptive evolution of the *G6pd* gene in *Drosophila melanogaster* and *Drosophila simulans* lineages. *Proc. Natl. Acad. Sci. USA* **90**: 7475–7479.
- GILLESPIE, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.
- HALL, J. G., 1985 The adaptation of allozymes to temperature: catalytic characterization of glucosylphosphate isomerase homologues isolated from *Mytilus edulis* and *Isognomon alatus*, bivalve molluscs inhabiting different thermal environments. *Mol. Biol. Evol.* **2**: 251–269.
- HARRIS, H., 1966 Enzyme polymorphisms in man. *Proc. R. Soc. Lond. B Biol. Sci.* **164**: 298–310.
- HARRISON, R. G., 1977a Parallel variation at enzyme locus in sibling species of field crickets. *Nature* **266**: 168–170.
- HARRISON, R. G., 1977b Patterns of variation and genetic differentiation in closely related species: the field crickets of eastern North America. Ph.D. Thesis, Cornell University, Ithaca, NY.
- HARRISON, R. G., 1979 Speciation in North American field crickets: evidence from electrophoretic comparisons. *Evolution* **33**: 1009–1023.
- HARRISON, R. G., and S. M. BOGDANOWICZ, 1995 Mitochondrial phylogeny of North American field crickets: perspectives on the evolution of life cycles, songs and habitat associations. *J. Evol. Biol.* **8**: 209–232.
- HEBERT, P. D. N., and M. J. BEATON, 1993 *Methodologies for Allozyme Analysis Using Cellulose Acetate Electrophoresis*. Helena Laboratories, Beaumont, TX.
- HOFFMAN, R. J., 1981 Evolutionary genetics of *Metridium senile*. I. Kinetic differences in phosphoglucose isomerase allozymes. *Biochem. Genet.* **19**: 129–144.
- HOFFMAN, R. J., 1983 Temperature modulation of the kinetics of phosphoglucose isomerase genetic variants from the sea anemone *Metridium senile*. *J. Exp. Zool.* **227**: 361–370.
- HUDSON, R. R., 1982 Estimating genetic variability with restriction endonucleases. *Genetics* **100**: 711–719.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, edited by D. FUTUYMA and J. ANTONOVICS. Oxford University Press, Oxford, England.
- HUDSON, R. R., M. KREITMAN and M. AGUAD, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- JOHANNESSON, K., N. KAUTSKY and M. TEDENGREN, 1990 Genotypic and phenotypic differences between Baltic and North Sea populations of *Mytilus edulis* evaluated through reciprocal transplantations. II. Genetic variation. *Mar. Ecol. Prog. Ser.* **59**: 211–219.
- KAROTAM, J., T. M. BOYCE and J. G. OAKESHOTT, 1995 Nucleotide variation at the hypervariable esterase 6 isozyme locus of *Drosophila simulans*. *Mol. Biol. Evol.* **12**: 113–122.
- KATZ, L. A., 1996 Molecular evolution of phosphoglucose isomerase at two phylogenetic levels. Ph.D. Thesis, Cornell University, Ithaca, NY.
- KATZ, L. A., 1997 Characterization of the phosphoglucose isomerase gene from crickets: an analysis of phylogeny, amino acid conservation and nucleotide composition. *Insect Mol. Biol.* **6**: 1–14.
- KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**: 412–417.
- LABATE, J., and W. F. EANES, 1992 Direct measurement of in-vivo flux differences between electrophoretic variants of G6PD from *Drosophila melanogaster*. *Genetics* **132**: 783–787.
- LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- LEWONTIN, R. C., and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**: 595–605.
- PATARNELLO, T., P. M. BISOL and B. BATTAGLIA, 1989 Studies on differential fitness of PGI genotypes with regard to temperature in *Gammarus insensibilis* (Crustacea: Amphipoda). *Mar. Biol.* **102**: 355–359.
- PLACE, A. R., and D. A. POWERS, 1979 Genetic variation and relative catalytic efficiencies: lactate dehydrogenase B allozymes of *Fundulus heteroclitus*. *Proc. Natl. Acad. Sci. USA* **76**: 2354–2356.
- POWERS, D. A., T. LAUERMAN, D. CRAWFORD and L. DIMICHELE, 1991 Genetic mechanisms for adapting to a changing environment. *Annu. Rev. Genet.* **25**: 629–659.
- RIDDOCH, B. J., 1993 The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). *Biol. J. Lin. Soc.* **50**: 1–17.
- RILEY, M. A., S. R. KAPLAN and M. VEUILLE, 1992 Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **9**: 56–69.
- SWOFFORD, D. L., 1995 PAUP, Test version 4.0d55.
- SYSTAT, 1996 SYSTAT for Windows: Statistics, Systat, Inc., Evanston, IL.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAYLOR, W. R., 1986 The classification of amino acid conservation. *J. Theor. Biol.* **119**: 205–218.
- TERAUCHI, R., T. TERACHI and N. T. MIYASHITA, 1995 DNA polymorphism and evolutionary history of the *Pgi* locus in *Dioscorea tokoro*, a wild yam species, pp. 245–248 in *Current Topic on Molecular Evolution*, edited by M. NEI and N. TAKAHATA. Inst. Mol. Evol. Genetics, Pennsylvania State University, State College, PA and Graduate School for Advanced Studies, Hayama, Japan.
- VEUILLE, M., and L. M. KING, 1995 Molecular basis of polymorphism at the esterase-5B locus in *Drosophila pseudoobscura*. *Genetics* **141**: 255–262.
- WATT, W. B. 1983 Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. *Genetics* **103**: 691–724.
- WATT, W. B., 1992 Eggs, enzymes and evolution: natural genetic

- variants change insect fecundity. *Proc. Natl. Acad. Sci. USA* **89**: 10608–10612.
- WATT, W. B., 1994 Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool? *Genetics* **136**: 11–16.
- WATT, W. B., R. C. CASSIN and M. S. SWAN, 1983 Adaptation at specific loci. III. Field behavior and survivorship differences among colias PGI genotypes are predictable from in vitro biochemistry. *Genetics* **103**: 725–739.
- WATT, W. B., P. A. CARTER and S. M. BLOWER, 1985 Adaptation at specific loci. IV. Differential mating success among glycolytic allozyme genotypes of *Colias* butterflies. *Genetics* **109**: 157–175.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- ZERA, A. J. 1987 Temperature-dependent kinetic variation among phosphoglucose isomerase allozymes from the wing-polymorphic water strider, *Limnoporus canaliculatus*. *Mol. Biol. Evol.* **4**: 266–286.

Communicating editor: A. G. CLARK