Extensive Amino Acid Polymorphism at the *Pgm* Locus Is Consistent With Adaptive Protein Evolution in *Drosophila melanogaster*

Brian C. Verrelli and Walter F. Eanes

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245

Manuscript received April 18, 2000
Accepted for publication July 31, 2000

ABSTRACT

PGM plays a central role in the glycolytic pathway at the branch point leading to glycogen metabolism and is highly polymorphic in allozyme studies of many species. We have characterized the nucleotide diversity across the *Pgm* gene in *Drosophila melanogaster* and *D. simulans* to investigate the role that protein polymorphism plays at this crucial metabolic branch point shared with several other enzymes. Although *D. melanogaster* and *D. simulans* share common allozyme mobility alleles, we find these allozymes are the result of many different amino acid changes at the nucleotide level. In addition, specific allozyme classes within species contain several amino acid changes, which may explain the absence of latitudinal clines for PGM allozyme alleles, the lack of association of PGM allozymes with the cosmopolitan *In(3L)P* inversion, and the failure to detect differences between PGM allozymes in functional studies. We find a significant excess of amino acid polymorphisms within *D. melanogaster* when compared to the complete absence of fixed replacements with *D. simulans*. There is also strong linkage disequilibrium across the 2354 bp of the *Pgm* locus, which may be explained by a specific amino acid haplotype that is high in frequency yet contains an excess of singleton polymorphisms. Like *G6pd*, *Pgm* shows strong evidence for a branch point enzyme that exhibits adaptive protein evolution.

UNDERSTANDING how selection at higher phenotypic levels impacts molecular variation at the single gene level is an important question in evolutionary and physiological genetics. A working view from a physiological perspective would propose that selection on life history variation results in selection to modulate energy budgets and this targets the partitioning of metabolic fluxes into growth and reproduction. Central to this reductionist view is how the differential fluxes at metabolic crossroads, such as those involving the glycolytic pathway and its numerous junctions, respond to selection. Historically, allozyme studies revealed much about polymorphism in these genes, and the observation that some enzymes are variable, while others are not, has potential adaptive explanations (Gillespie 1991; Mitton 1998). A fundamentally important question is whether intrinsic features of these enzymes, such as their structure-function constraints or relative roles in partitioning fluxes among competing pathways, lead to the gene-specific patterns of intra- and interspecific variation. Studies of *G6pd*, which shares a common substrate with both phosphoglucomutase and hexokinase at the head of the glycolytic pathway at the branch leading to the pentose shunt, make a compelling case for both selection acting on the enzyme polymorphism in *Drosophila melanogaster* and episodic selection on amino acid changes in the lineage leading to *D. simulans* (Eanes *et al.* 1993, 1996). To understand this problem in a larger context, a systematic study of the enzymes of the pathway is important (Eanes 1999). This lab has focused on the need to characterize molecular variation in these genes and in this report addresses the nature of molecular variation found at the *Pgm* locus in *D. melanogaster* and *D. simulans*.

Phosphoglucomutase (PGM; EC 2.7.5.1) plays a major role in the synthesis and breakdown of glycogen, which is important to energy storage in muscle tissue, and catalyzes the interconversion of glucose-1-phosphate and glucose-6-phosphate (Ray and Roscelli 1964; Hirose *et al.* 1970). PGM shares glucose-6-phosphate with several other enzymes at a branch point at the head of the metabolic pathway leading into glycogen metabolism, the pentose shunt, and the main glycolytic corridor. PGM has been routinely studied in many organisms because of its high level of allozyme polymorphism (Dawson and Jaegar 1970), and there are three common allozymes in *D. melanogaster* (Hjorth 1970; Trippa *et al.* 1970). The *Medium* mobility allele is widespread, but seven additional allozyme alleles and other underlying amino acid variants have been revealed by electrophoretic and thermostability criteria (Trippa *et al.* 1976, 1977, 1978). A study of geographic variation of the allozyme alleles found no evidence of clines (Oakeshott *et al.* 1981), which differs from other allozyme polymorphisms in *D. melanogaster*, where latitudinal clines are common and taken as supporting evidence for the action of natural selection (Singh and Rhomberg 1987;

Corresponding author: Brian C. Verrelli, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245. E-mail: verrelli@life.bio.sunysb.edu

Eanes 1999). *D. simulans* also has three common allozymes possessing the same electrophoretic mobilities and geographic patterns as the three common allozymes in *D. melanogaster*. This raises the interesting possibility that the common PGM allozyme polymorphisms could predate speciation (Anderson and Oakeshott 1984; Hyytia *et al.* 1985), and in a pair of species for which very few shared polymorphisms exist, a locus with shared allozyme alleles suggests balancing selection maintains this protein polymorphism. In addition, several lab selection experiments also suggest balancing selection influences PGM allozyme allele frequencies, although other studies show inconsistent patterns (Carfagna *et al.* 1980; Oakeshott *et al.* 1985, 1988). A biochemical study of the allozyme alleles in *D. melanogaster* reported no difference in kinetic properties (Fucci *et al.* 1979). Nonetheless, if a specific allozyme mobility class results from several different amino acid differences (as suggested by the thermostability studies), this will potentially confound both studies of geographic variation and functional studies where variation is defined only by allozyme mobility. Finally, the *Pgm* locus (located on the third chromosome at 72D7) also lies just inside the proximal breakpoint (73E3) of the cosmopolitan *In(3L)P* inversion. This inversion exhibits clines with latitude in natural populations on several continents (Mettler *et al.* 1977; Knibb *et al.* 1981), and, in principle, selection favoring the inversion could actually maintain variation at the *Pgm* locus (Wesley and Eanes 1994; Hasson and Eanes 1996). Despite this tight physical linkage, no significant linkage disequilibrium has been found between PGM allozyme alleles and the inversion in natural populations (Langley *et al.* 1974, 1977).

Given our interest in examining patterns of variation in metabolic genes, we were interested in addressing several questions in this study. We first recover and characterize the primary sequence for *Pgm* from *D. melanogaster* and from this investigate overall sequence variation and the molecular nature of the allozyme variation. We address whether *Pgm* is similar to *Est-6* in exhibiting abundant amino acid variation (as predicted from thermostability studies), as well as possessing high levels of interspecific divergence. Because *D. melanogaster* and *D. simulans* share the same allozyme mobility alleles for PGM, we will also examine the possibility of a common mutational origin and long-term persistence of these allozymes in both species. Finally, we were interested in determining whether there is evidence for adaptive protein evolution at the *Pgm* locus like that seen for other metabolic enzymes such as *G6pd* (Eanes *et al.* 1993) and *Adh* (McDonald and Kreitman 1991).

MATERIALS AND METHODS

Isolation of *Pgm* locus: A third instar larvae cDNA library made from *D. melanogaster* was kindly provided by Peter Gergen (SUNY at Stony Brook). Yeast and human PGM protein sequences (GenBank accession nos. P33401 and P36871, respectively) were aligned and four degenerate primers (two positive, two negative) were constructed from highly conserved regions of the sequence. Using these primers, a 300-bp fragment was amplified from the cDNA library via PCR. This fragment was sequenced with the degenerate primers to verify the isolation of *D. melanogaster Pgm* cDNA. Primers specific to the *D. melanogaster* sequence were then constructed from the 300-bp fragment. The cDNA library was then amplified with vector-specific primers and the cDNA pool was size selected from a 3% agarose gel to remove other abundant transcripts of unwanted sizes. The recovered 300-bp *Pgm* fragment was labeled with biotin-14-dATP using PCR, denatured at 95° for 5 min, and hybridized to the size-selected single-stranded cDNA pool at 45° for 8 hr. This DNA-duplex was then isolated with streptavidin-coated magnetic beads (Dynal, Inc., Great Neck, NY) at room temperature for 20 min. After several washes to remove unspecific fragments bound to the beads, the DNA was eluted from the magnetic beads in 2.5 mm EDTA at 85°, and this template DNA was used directly for PCR. A specific *Pgm* positive primer was paired with a negative cDNA primer to obtain the remaining sequence to the 3′ end. However, due to the reverse transcriptase enzyme prematurely falling off during transcription, a heterogeneous fragment pool was amplified when a positive vector-specific primer was paired with a negative *Pgm* specific primer. This PCR fragment pool was cloned and 100 clones were subsequently used for PCR templates. The largest cloned insert from this screen was sequenced and found to be short of the 5′ end. A new negative *Pgm* specific primer was synthesized from this sequence and combined with the positive cDNA vector-specific primer to amplify another pool of *Pgm* transcripts from the template DNA recovered from the magnetic beads. This "transcript walk" of amplification, cloning, sequencing, and primer synthesis was performed several times until the 5′ end of the coding region was finally discovered. A new positive *Pgm* specific primer was then synthesized at the 5′ end to amplify the entire gene from genomic DNA. Introns were determined when full genomic sequence was compared to our previously recovered *Pgm* cDNA sequence.

Origin of wild lines: *D. melanogaster* isofemale lines were collected from a Davis Peach Farm, Mt. Sinai, New York, population in 1995 (DPF95) and made homozygous for the third chromosome using the *TM3/TM6* balancer. Lines were genotyped for allozyme mobility (Hjorth 1970), and based on the frequency of the three common allozyme mobilities (*Medium*, *Fast*, and *Slow*) in this population, a constructed random sample (CRS, Hudson *et al.* 1994) of 22 lines was sequenced. Therefore, parameter estimates (*i.e.*, θ and π) calculated from this sample reflect the frequencies of PGM allozyme alleles in this population. Isofemale lines collected in 1990 from two populations in Zimbabwe [Havare, Z(H) and Sengwa Wildlife Preserve, Z(S)] were kindly provided by C.-I. Wu's laboratory. Thirteen of these lines were made homozygous for the third chromosome and were sequenced for *Pgm* to compare the level of variation in Zimbabwe with the North American population sample. Using primers specific for the *In(3L)P* arrangement (Wesley and Eanes 1994), extracted third chromosome lines were screened for the inversion and *Pgm* was sequenced from these: Davis Peach Farm, Mt. Sinai, New York (3 lines; DPF95), Maryland (1 line; MD90), Spartanburg, South Carolina (1 line; SC96), and Homestead, Florida (5 lines; HFL97). *D. simulans* isofemale lines were collected from a Davis Peach Farm population in 1996 (DPF96), and a survey for the three PGM allozyme allele frequencies was performed. We sequenced a CRS of 13 lines, which reflects the frequencies of the alloyzme alleles in this population. Lines genotyped as a *Fast* or a *Slow* allozyme mobility were made homozygous at

the *Pgm* locus after two generations of mating with a *Medium* mobility allele stock. Conversely, lines genotyped with a *Medium* allozyme mobility were made homozygous at the *Pgm* locus after two generations of mating with a *Fast* mobility allele stock. Flies were then cut in half, and the abdomen was used to check allozyme genotype while the thorax served as tissue for DNA extraction. Nine *D. yakuba* isofemale lines from a 1999 West African sample (T. F. C. Mackay's laboratory) were sequenced and used to partition fixations into the *D. melanogaster* and *D. simulans* lineages. Finally, a PGM null allele (line Pgm^{nGB1}) found in a natural population sample from Great Britain (Langley *et al.* 1981; Burkhart *et al.* 1984) was obtained from the Bloomington Stock Center (B4039) and sequenced.

PCR amplification and sequencing: Fragments were amplified in 10-μl volumes in an Idaho Technologies (Idaho Falls, ID) Air-Thermo-Cycler by PCR from single-fly CTAB genomic preps (Winnepenninckx *et al.* 1993). PCR products are excised from 2% agarose gels for 50-μl reamplifications using internal primers. DNA fragments were purified from PCR reactions [Bio-Rad (Hercules, CA) Prep-A-Gene kit] and double-stranded DNA templates were manually sequenced using the Sequenase kit (United States Biochemical Co., Cleveland) and [^{35}S]dATP (Amersham, Buckinghamshire, England). Reactions were run on acrylamide gels with an electrolyte gradient and electrophoresed for 3–7 hr. All sequences are stored under GenBank accession nos. AF290313–AF290370.

RESULTS

Levels of intraspecific polymorphism: The isolation of the primary *Pgm* sequence from the *D. melanogaster* cDNA library yielded a 1680-bp coding region equal to 560 amino acids in length and shows high similarity to known PGM protein sequences from many other taxa (Whitehouse *et al.* 1998). Upon sequencing genomic DNA, four exons of 78, 247, 1219, and 136 bp and three introns of 71, 534, and 66 bp were discovered. There are 408 silent site equivalents across the 1680-bp coding region. The CRS of 22 lines (DPF95) is comprised of 3 *Slow*, 2 *Fast*, and 17 *Medium* allozyme alleles and contains 13 silent, 12 replacement, and 17 intron polymorphisms. Figure 1 lists a total of 55 variable sites from all 44 lines sequenced for *D. melanogaster*. The three *Slow* allozyme alleles in this data set result from three independent amino acid replacements at nucleotide sites 25 (Ala to Thr at amino acid residue 9, or A9T), 1194 (E197K), and 1308 (E235K). Although the latter two replacements explain the charge changes of the slower electrophoretic mobilities, the first replacement at nucleotide site 25 (A9T) does not predict a charge change. Additional sequence data for the *Slow* allozyme mobility class reveals that this class is truly heterogeneous; however, the polymorphism at nucleotide site 25 accounts for most of the *Slow* copies in natural populations (our unpublished data). The two *Fast* allozyme alleles in the CRS both possess a pair of tightly linked changes at nucleotide sites 1324 (R240L) and 1340 (E245D), the change at nucleotide site 1324 being responsible for the *Fast* electrophoretic mobility. The remaining 8 replacement polymorphisms in Figure 1 are cryptic amino acid changes that are not detectable in allozyme screens. The *D. simulans* sequences predict that the *Medium* allele is the ancestral state. The average numbers of silent differences within the *Medium*, *Slow*, and *Fast* allele classes are 1.76, 4.00, and 0.00. On average, *Medium* and *Fast* alleles differ by 2.35 silent sites, *Medium* and *Slow* alleles differ by 2.74 silent sites, and *Fast* and *Slow* alleles differ by 2.67 silent sites.

Sequence variation for the Davis Peach Farm CRS, the Zimbabwe sample, and the *In(3L)P* sample are all summarized in Table 1. The 13 sequences from Zimbabwe add an additional five silent, four replacement, and two intron polymorphisms to the North American sample in Figure 1. A single *Slow* allozyme allele was sequenced in this sample (Z26H) and shares the same change at nucleotide site 25 as that found in the North American sample (DPF95 94.1). The level of silent polymorphism in the Zimbabwe sample is actually lower than that of the North American sample. This is not a typical observation seen in comparisons made between these two geographic regions (Begun and Aquadro 1993; Eanes *et al.* 1996). The Zimbabwe sample segregates several of the North American replacements in addition to a few unique singletons. The CRS of 22 lines was screened for the *In(3L)P* inversion and two lines, DPF95 13.0 and DPF95 48.2, both bear the inversion. Eight additional lines containing the inversion are also listed in Figure 1 and together add only one unique silent polymorphism. A unique mutation (site 1308) resulting in a *Slow* allele (DPF95 13.0) is found on an inverted chromosome only. A surprising observation, given the proximity of the breakpoint, is the lack of divergence between *Pgm* alleles on standard and inverted arrangements. Figure 1 shows the presence of shared polymorphism between arrangements and indicates some level of exchange. A single gene conversion event is predicted between the two arrangements at nucleotide sites 226–460 (235 bp in length) using DnaSP (Rozas and Rozas 1999), and exchange of short "gene tracts" of this nature alone may explain the lack of divergence between the two arrangements. It seems that although most of the observed variation is shared among the three *D. melanogaster* samples in Figure 1, all three samples contain private alleles as well. Nucleotide divergence between the Davis Peach Farm CRS and the Zimbabwe sample is 0.0054 and 0.0015 at silent and replacement sites, respectively. Nucleotide divergence between the Davis Peach Farm CRS and the *In(3L)P* sample is 0.0059 and 0.0016 at silent and replacement sites, respectively. Overall, these three samples are not apparently different in silent or replacement variation and divergence at the *Pgm* locus. Finally, the Pgm^{nGB1} null allele has a unique replacement polymorphism at nucleotide site 931 (G109A) and this change lies in the active site of PGM (Dai *et al.* 1992; Liu *et al.* 1997). This glycine to alanine substitution is not a conservative biochemical change (Argos *et al.* 1979; Chakrabartty *et al.* 1991)

```
                                    Nucleotide position
                                      11111111111111112 2222 2222222
                   1 22222223 4455566667889 911333355666788990 1112 2222333
               257 90 02226771 6825847781022 359024937124426795 5790 1589224
               518 87 00562287 0337404869615 174840421762262385 6854 6997178
Line           a                               aab
SIM            GGG GG CCGTTGTC TGGGGGCTAT-TT GGGGGACGCGGGCAGCAG CTCG CGAGCAC
DPF95 3.0      ... .. ...C.... .........AT.. .................. .CA. .......
DPF95 23.1     ... .. ...C.... .........AT.. .................. .CA. .......
DPF95 84.3     ... .. ...C.... .........AT.. .................. .CA. .......
DPF95 4.3      ... .. ...C.... .A.......AT.. .................. .CA. .......
DPF95 29.3     ... .. ...C.... .........AT.. ................T. .CA. .......
DPF95 36.4     ... .. ...C.... .........AT.. .........T........ .CA. .......
SC96 5.3*      ... .. ...C.... .........AT.. ..............T... .CA. .......
DPF95 56.2*    ... .. ...C.... .........AT.. ..............T... .CA. .......
DPF95 13.0*    ... .. ...C.... .........AT.. ...A..........T... .CA. .......
DPF95 54.0     ... .. ...C.... .....A...AT.. .................. .CA. .......
DPF95 2.1      ... .. ...C.... .....A...AT.. .................. .CA. .......
DPF95 73.1     ... .. ...C.... .....A...AT.. .................. .CA. ....T..
DPF95 77.4     ... .. ...C.... .....A...AT.. ..........A....... .CA. .......
DPF95 90.2     ... .. ...C.... .........AT.. ..........AA...... .CA. .......
Pgm^nGB1       ... .. ...C.... .........AT.. C.........A....... .CA. .......
DPF95 94.1     A.. .C ...C.... ...AA.A..AT.. ...............A.. .CA. ...A...
Z(H) 26        A.. .. ...C.... ...AA.A..AT.. .................. .CA. .......
DPF95 4.2      ... .. T..C.... ..C..A.A..TC. ....TT...........T A.A. .....C.
DPF95 38.3     ... .. T..C.... ..C..A.A..TC. ....TT...........T A.A. .....C.
Z(S) 49        ... .. T..C.... ..C..A.A..TC. .................T A.A. .....CT
DPF95 44.3     .A. .. T..C..A. ..C..A.A..TC. .................T A.A. .A...C.
DPF95 85.1     .A. .. T..C..A. ..C..A.A..TC. .................T A.A. .A...C.
Z(H) 23        ... .. T..C.... ..C....A..TC. .................T A.A. .....C.
Z(S) 11        ... .. T..C.... ..C....A..TC. .................T A.A. .....C.
Z(S) 48        ... .. T..C.... ..C....A..TC. ............T....T A.A. ..T..C.
DPF95 56.1     ... .. T..C.... ..C.......TCC .AA..............T A.A. .....C.
Z(H) 44        ... .. T..C.... ..C.......T.. .................T A.A. .....C.
Z(H) 38        ... .. TT.C.... ..C....A..TC. .................T ..A. .......
Z(S) 24        ... .. T..C.... ..C....A..TC. .................T ..A. .......
Z(S) 35        ... .. T..C.... ..C....A..TC. ........T........T .... .......
MD90 709.1*    ... .. ...C.... ..........T.. ......T....A.....T A.A. .....C.
HFL97 1.0*     ... .. ...C.... ..........T.. ......T....A.....T A.A. .....C.
HFL97 13.0*    ... .. ...C.... ..........T.. ......T....A.....T A.A. .....C.
HFL97 15.0*    ... .. ...C.... ..........T.. ......T....A.....T A.A. .....C.
HFL97 50.0*    ... .. ...C.... ..........T.. ......T....A.....T A.A. .....C.
DPF95 53.1     ..A .. ...C.... .........AT.. .............G...T A.A. .A...C.
Z(H) 39        ... A. ...C.A.. .......A..T.. ...........A.....T ..A. .......
Z(S) 51        ... .. ...C.... ..C.......T.. .......T.......... ...A .......
Z(S) 15        ... .. ...C.... ..........T.. .......T.......... ...A .......
DPF95 2.0      ... .. ...C.... ..........T.. .................. .... .......
DPF95 48.2*    ... .. .......T C.........A.. .................T A... A......
HFL97 93.0*    ... .. .......T C.........A.. .................T ..A. .......
DPF95 100.3    ... .. .......T C.......G.T.. .............G.... .... .......
Z(H) 36        ... .. ..A.A..T ..C.......T.. ...........A.....T ..A. .......
```

Figure 1.—List of 55 polymorphic sites identified across 2354 bp of *Pgm* in 44 lines of *D. melanogaster*. Characters **a** and **b** designate nucleotide positions responsible for *Slow* and *Fast* allozyme polymorphisms, respectively. Nucleotide polymorphisms shown in boldface type are amino acid polymorphisms. Italics refer to polymorphisms in introns I, II, and III, respectively. SIM refers to *D. simulans* line DPF96 3.0. Line labels followed by an asterisk have the *In(3L)P* inversion.

and may explain the reported loss of PGM activity for this allele.

The summary statistics for the 13 *D. simulans* alleles are presented in Tables 1 and 2. The CRS of 2 *Slow*, 1 *Fast*, and 10 *Medium* alleles are presented in Figure 2. When compared to *D. melanogaster*, Table 1 shows almost a sixfold increase in the level of silent site polymorphism for *D. simulans* with 64 silent site polymorphisms in the sample. The level of replacement polymorphism is lower, which is typical for these two species (Moriyama and Powell 1996). Five replacement and 22 intron polymorphisms are also shown in Figure 2. All 5 of the replacement polymorphisms result in rare allozyme alleles, none of which are shared with *D. melanogaster*. The two sampled *Slow* allozyme alleles are the result of two different replacements at the nucleotide level. Line DPF96 10.6 has a second replacement polymorphism at nucleotide site 217 (A49G) in addition to the change responsible for its *Slow* mobility at site 205 (E45A). Line DPF96 17.4 has a change at nucleotide site 1128 (D175N) that results in the second *Slow* allele. Line DPF96 36.2 has a second replacement polymorphism at nucleotide site 1954 (T450I) in addition to the change responsible for its *Fast* mobility at nucleotide site 1075 (K157M). Additional *Fast* alleles were sequenced (our unpublished data), and all show the same polymorphism at nucleotide site 1075. In contrast to *D. melanogaster*, the second intron is highly variable in length within *D. simulans* and is omitted from the data in Figure 2. An apparent 2-bp insertion in the third intron in *D. melanogaster* and the highly variable second intron in *D. simulans* account for the only differences in gene length for the two species.

Figure 3 shows the neighbor-joining analysis (Saitou and Nei 1987) of all 44 sequences of *D. melanogaster* and 13 sequences of *D. simulans*. It is based only on silent and intron polymorphisms within and between both species, and this illustrates the clustering of silent varia-

TABLE 1

Summary parameter estimates for *D. melanogaster* and *D. simulans*

Sample	n	θ_s	θ_a	π_s	π_a
D. melanogaster					
CRS	22	0.0087	0.0026	0.0052	0.0015
Zimbabwe	13	0.0063	0.0020	0.0051	0.0013
In(3L)P	10	0.0043	0.0011	0.0052	0.0012
D. simulans	13	0.0506	0.0013	0.0483	0.0007

CRS refers to the constructed random sample (see text) of 22 *D. melanogaster* sequences from the Davis Peach Farm population. All estimates are based on 408 silent site equivalents. θ_s and π_s and θ_a and π_a are measures of nucleotide diversity for silent and replacement sites, respectively.

tion independent of the amino acid polymorphisms. Although recombination can obscure the true genealogical relationship among allele copies, this analysis is used simply as an exploratory method to structure the variation at this locus. However, the analysis using all polymorphic sites (including amino acid polymorphisms) results in the same clustering of alleles indicating strong disequilibrium across the gene. As previously pointed out in Figure 1, Figure 3 demonstrates that there is no evidence of geographic structuring when the Davis Peach Farm and Zimbabwe populations are compared in a genealogical framework. In addition, the *In(3L)P* sequences do not form one single cluster, but three small clusters, representing gene conversion events that have led to some divergence between alleles on inverted chromosomes and convergence with alleles on the standard arrangements. Two *In(3L)P* sequences branch in a cluster that is ancestral to all other sequences, including all other standard alleles, an observation not uncommon for this inversion (Hasson and Eanes 1996). All allozyme alleles are also labeled in Figure 3. Although several *Pgm* alleles share electrophoretic mobilities, both species possess *Slow* mobility alleles that are more closely related to *Medium* and *Fast* alleles than they are to other *Slow* alleles. The difference in levels of silent polymorphism within each species is reflected in the deeper branches in *D. simulans* compared to *D. melanogaster*.

Linkage disequilibrium: There is pervasive linkage disequilibrium across the entire 2354-bp region of the *Pgm* gene in *D. melanogaster*. This is notable from Figure 4 in the high associations of polymorphisms for the 22 sequences of the Davis Peach Farm CRS. Using 21 polymorphic sites (21 singletons are omitted), 99 out of 210 pairwise correlations are statistically significant by a chi-square test ($P < 0.05$). Figure 4 also shows that nucleotide sites more than 2 kb apart show some of the strongest disequilibria (33 of the 210 pairwise correlations are significant at the 0.1% level with a Bonferroni correction). Because two of the sequences in our Davis Peach Farm CRS (DPF95 48.2 and DPF95 13.0) are associated with the *In(3L)P* inversion, this could inflate estimates of linkage disequilibrium for the entire data set. However, because there is no apparent association between the inversion and *Pgm* variation, and because no divergence exists between arrangements, linkage disequilibrium estimates did not change even when these two alleles were removed from the analysis.

Tests of the polymorphism frequency spectrum: Tajima (1989) and Fu and Li (1993) tests were applied to the data for both species. We tested the Davis Peach Farm CRS of 22 sequences of *D. melanogaster*, the 13 Zimbabwe sequences, the 10 *In(3L)P* sequences, and the 13 *D. simulans* sequences, each independently, to see if either the silent or replacement polymorphism frequency distributions differed from that expected for a neutral distribution of variation. All test results are presented in Table 3. Only the analysis of *D. simulans* replacement polymorphism by the Fu and Li test is significant.

TABLE 2

Intra- and interspecific comparisons for *D. melanogaster* and *D. simulans* for the 1680-bp coding region of *Pgm*

	Silent	Replacement	Fisher exact test
D. melanogaster and *D. simulans*			
Polymorphic	77	17	
Fixed	34	0	$P < 0.001$
D. melanogaster			
Polymorphic	13	12	
Fixed	19	0	$P < 0.0001$
D. simulans			
Polymorphic	64	5	
Fixed	11	0	$P = 0.63$

Silent site divergence is partitioned into the two lineages by use of *D. yakuba* data and tested *vs.* silent site polymorphism within each lineage. The total number of fixed differences does not equal the sum of the fixed differences in the two lineages because several fixations were ambiguous when compared with *D. yakuba*.

Line	Nucleotide position			
			111111111111111111	
		11111	122222223333344	000011111122223333
	2	*889999900024*	700112990244800	677812236623480133
	7	*230456814672*	335577392417068	459921803902491747
			a	**b** **a**
MEL			C AGAGGCGACCCA	CG**AA**CTGACTCCTAG
DPF96 7.0			. .....A.....	.........C.....
DPF96 20.3			. .....A.....	.........C.....
DPF96 18.0			. .....A.....	.........C.....
DPF96 9.4			. .....A.....	.........C.....
DPF96 10.6			. ....AA.....C	..**C.G**..CT..T..A
DPF96 36.2			. TAG..A...G.G	...T.....C...C.
DPF96 13.1			. .....A.....	.....A.....
DPF96 13.2			. .AG..A..T...	.....A...CT.G..
DPF96 3.0			. .....A..T...	.....A....T.G..
DPF96 22.2			. .....A..T...	.....A....T.G..
DPF96 17.4			T .....G.....	T..T..A.....A
DPF96 5.1			. ...A.AAG..T.	.A...A.....
DPF96 43.3			. ...A.AAG..T.	.A...A...C.....

Line	Nucleotide position			
		11111111111111111111111111111122222	*2222222222*	22
		33444445556666666677888999999900011	*1111111222*	23
		46112380150145677908369013579369O4	*6677899001*	82
		31297618764962109407822407401666 84	*1706656673*	57
MEL		ATCCCCCGAGCAGCGTATCGCGTCCC**C**CCAACGC	TTAATCGTCC	CA
DPF96 7.0		C..T.......CA....CTA.....A.....	.A.GA.A...	A.
DPF96 20.3		C..........CA....CTA.....A.....	.A.GA.A...	A.
DPF96 18.0		C..........CA....CTA.....A.....	.A.GA.A...	A.
DPF96 9.4		C..........CA....CTA.....A.....T	.A.GA.A...	A.
DPF96 10.6		C..TTT..G...AT...CTA.C.........T	.....C..	.C
DPF96 36.2		CC..T...G...AT.ACCTAT..TA.**T**.....	.....T.C..	..
DPF96 13.1		C.............CTA.....AT...AT	.....G.	..
DPF96 13.2		C.....T...T.....CTA.....A.GG...	C.G.....	..
DPF96 3.0		C.....T...T.....CT.....A...GG...	C.....	..
DPF96 22.2		C.....T...T.....CTA.....A...GG...	C.....	..
DPF96 17.4		C.............A....A..AT.....	.....	..
DPF96 5.1		..T....C.A.....C.A..C....A...T.T	...G.....T	..
DPF96 43.3		..T....C.A....T..C.A..C....A...T.T	...G.....T	..

FIGURE 2.—List of 92 polymorphic sites identified across 1852 bp of *Pgm* in 13 lines of *D. simulans*. Characters **a** and **b** designate nucleotide positions responsible for *Slow* and *Fast* allozyme polymorphisms, respectively. Nucleotide polymorphisms shown in boldface type are amino acid polymorphisms. Italics refer to polymorphisms in introns I and III, respectively (intron II omitted). MEL refers to *D. melanogaster* line DPF95 2.0.

Interspecific comparisons: Table 2 shows the summary of polymorphism and divergence for the total coding region of 1680 bp in *D. melanogaster* and *D. simulans*. Pooled polymorphisms for the two species result in 77 silent and 17 replacement polymorphisms. Comparing the Davis Peach Farm CRS of 22 sequences of *D. melanogaster* with the 13 sequences of *D. simulans*, the total numbers of fixed silent and replacement sites between the two species are 34 and 0, respectively. Because intraspecific polymorphism and interspecific divergence are correlated, a statistical test of independence can determine if these values deviate from that expected under neutrality (McDONALD and KREITMAN 1991). The 2 × 2 contingency Fisher's exact test of the pooled polymorphism (77/17) *vs.* the divergence at silent and replacement sites (34/0) is highly significant ($P < 0.001$). Using *D. yakuba* as an outgroup, the number of fixed differences along the *D. melanogaster-simulans* lineage can be partitioned into either species' lineage, as shown in Table 2. There is no significant difference in the number of silent fixations in either lineage. However, when the number of silent and replacement polymorphisms in each species is contrasted with the number of fixed silent and replacement differences by lineage using a Fisher's exact test, the *D. melanogaster* comparison is highly significant ($P < 0.0001$), whereas the *D. simulans* contrast is not. Compared to the level of divergence at silent and replacement sites along the *D. melanogaster* lineage (19/0), there appears to be a significant deviation in the ratio of silent to replacement polymorphism (13/12). The level of silent site polymorphism at *Pgm* ($\theta = 0.0087$) is typical of other *D. melanogaster* loci (MORIYAMA and POWELL 1996). Assuming that the silent site variation is neutral, the difference in the ratios of polymorphism to divergence at replacement and silent sites could indicate unusually high amino acid polymorphism or unusually low amino acid divergence. Because interspecific comparisons often show no replacement site divergence for these species (MORIYAMA and POWELL 1996), this suggests that the observation of high replacement polymorphism at the *Pgm* locus in *D. melanogaster* is unusual.

DISCUSSION

Excess of amino acid polymorphism: The principal observation of this study is the large number of amino

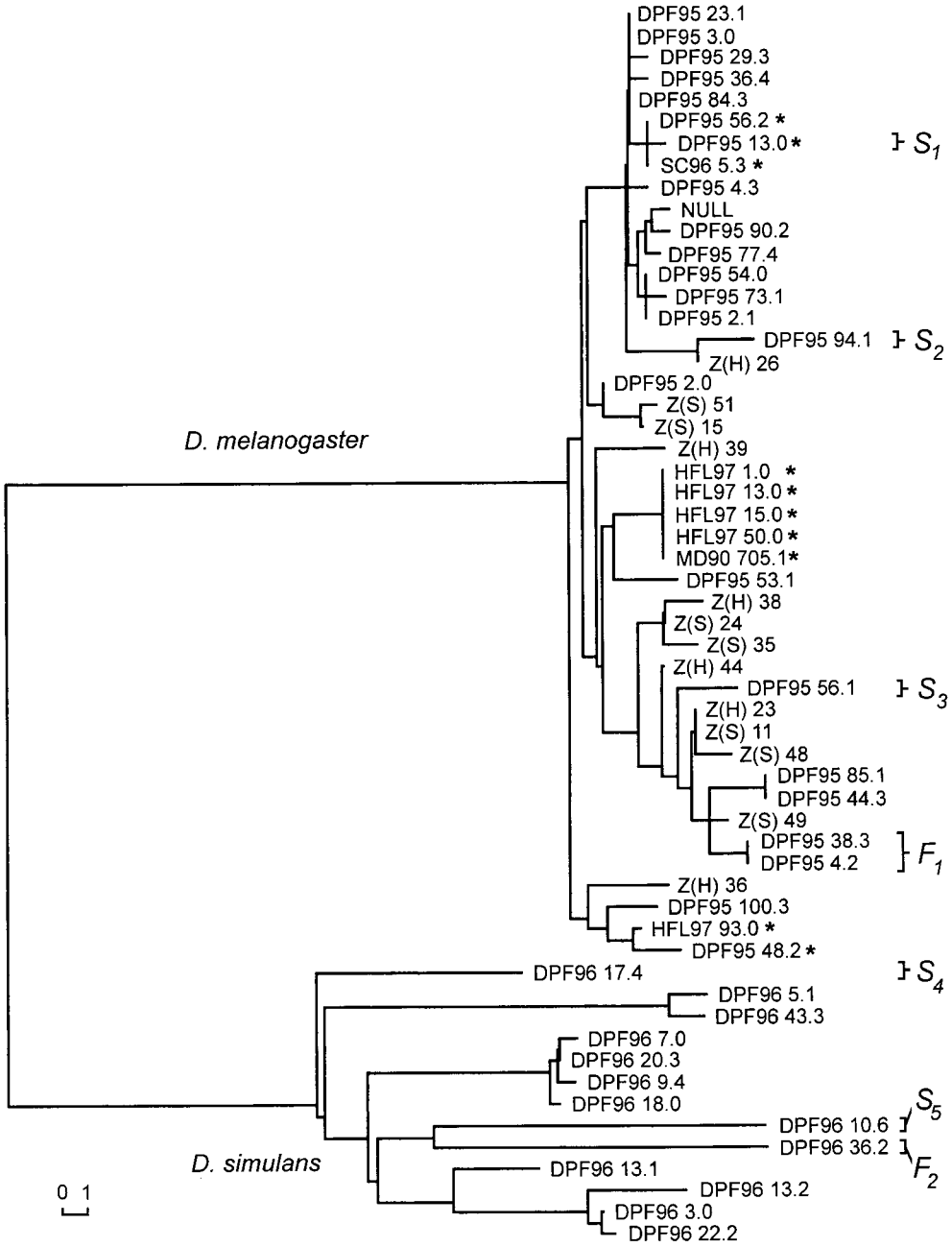

FIGURE 3.—Neighbor-joining tree based on silent substitutions among and between the 44 *D. melanogaster* and 13 *D. simulans Pgm* sequences. S_1, S_2, and S_3 designate *Slow* alloyzme alleles due to three unique mutations at the nucleotide level in *D. melanogaster*. S_4 and S_5 designate *Slow* allozyme alleles due to two unique mutations at the nucleotide level in *D. simulans*. F_1 and F_2 designate *Fast* allozyme alleles due to two unique mutations at the nucleotide level in *D. melanogaster* and *D. simulans*, respectively. Asterisks designate sequences with the *In(3L)P* inversion. Scale indicates one difference.

acid polymorphisms segregating in both species. The significance of this observation is supported by the McDonald-Kreitman test using the pooled data for both species. The overall level of silent divergence (8%) is typical of these two species (MORIYAMA and POWELL 1996) and therefore cannot explain the significant deviation in the McDonald-Kreitman test. By using *D. yakuba* as an outgroup, fixations along the *D. melanogaster-simulans* lineage can be partitioned into each species' lineage and compared to levels of silent and replacement polymorphism. Although the ratio of silent to replacement polymorphisms is near parity within the *D. melanogaster* lineage (13/12), the ratio for fixations at silent and replacement sites is dramatically different (19/0). Because the contrast of silent and replacement polymorphism (64/5) to silent and replacement divergence (11/0) in the *D. simulans* lineage is not statistically significant, the result from the pooled data from both species ($P < 0.001$) appears largely due to the significant excess of replacement polymorphism in the *D. melanogaster* lineage ($P < 0.0001$).

The relative difference between levels of polymorphism and divergence at the *Pgm* locus is atypical of other genes in *D. melanogaster*. Figure 5 plots the relationship of replacement polymorphism (characterized as θ_a) to divergence at replacement sites (d_a) for 17 genes. Sequence data for genes that exhibit protein polymorphism were taken from GenBank (see Figure 5

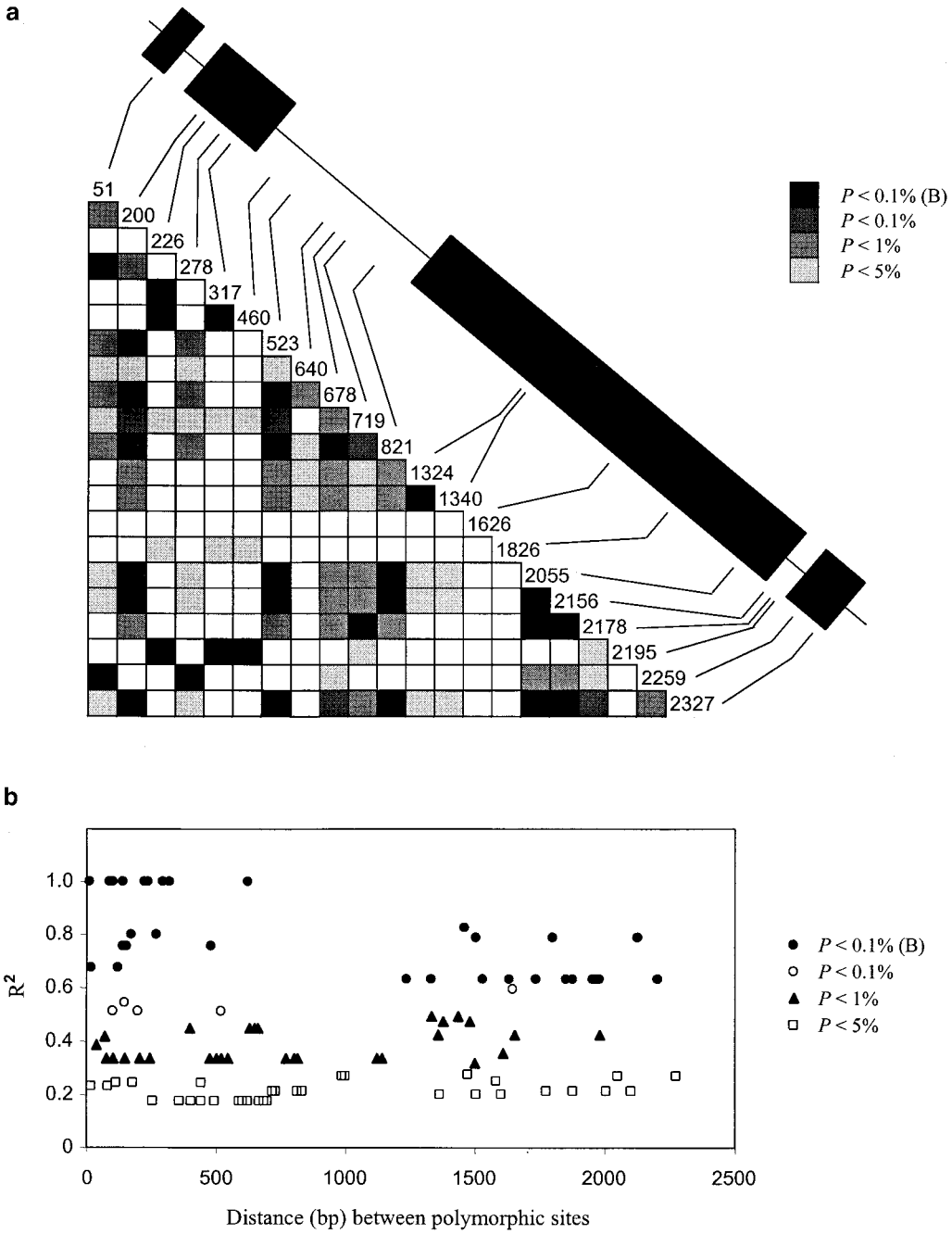

FIGURE 4.—(a) Shaded boxes indicate the strength of linkage disequilibrium across the 2354-bp gene of *Pgm* and designate the 99 of possible 210 associations between the 21 polymorphic sites (not including singletons) in both introns and exons that are significant by a chi-square test. The gene diagram above the correlation matrix displays the structure of the four exons (black boxes) and three introns for *Pgm*. The (B) in both plots refers to the 33 correlations that are significant at the 0.1% level with a Bonferroni correction. (b) Relationship between the measure of linkage disequilibrium R^2 and the distance between the 21 compared polymorphic sites.

for references for original data). WATTERSON's θ_a (1975) was calculated using the effective number of replacement sites, and divergence was calculated as the total number of fixed differences between species (compared with *D. simulans*) divided by the effective number of replacement sites. Overall there is a strong positive correlation ($r^2 = 0.587$; $P < 0.001$), and loci that show high levels of amino acid polymorphism also display high levels of amino acid divergence. An extreme example is *Est-6*, where a high level of amino acid polymorphism is paralleled by extensive amino acid divergence (COOKE and OAKESHOTT 1989; KAROTAM *et al.* 1995). This correlation is expected under the neutral theory because polymorphism is simply an intermediate stage in divergence between species. When the McDonald-Kreitman test is performed on each of the 17 genes for *D. melanogaster*, only 2 reject the neutral expectation. These outliers, *Pgm* and *G6pd*, both show highly significant departures, but in opposite directions. In addition, the exclusion of these two outliers from the plot in Figure 5 results in an even stronger correlation ($r^2 = 0.750$; $P < 0.0001$). *G6pd* possesses two protein polymorphisms, but displays a very high number of fixed replacements (EANES *et al.* 1993). This pattern is consistent with adaptive protein evolution where amino acid replacements are favored, fix rapidly, and are generally

TABLE 3

Neutrality tests for *D. melanogaster* and *D. simulans* intraspecific polymorphism

		Tajima's *D*		Fu and Li's *D*	
Sample	*n*	Silent	Rep	Silent	Rep
D. melanogaster					
CRS	22	−1.45	−1.47	−1.57	−1.32
Zimbabwe	13	−0.69	−1.39	−1.12	−1.41
In(3L)P	10	0.80	0.32	0.60	0.38
D. simulans	13	−0.13	−1.59	0.16	−2.74**

CRS refers to the constructed random sample of 22 *D. melanogaster* sequences from the Davis Peach Farm population. Tajima's *D* and Fu and Li's *D* are shown for tests of the silent and replacement polymorphism frequency spectrums for all four samples. $^{**}P < 0.01$.

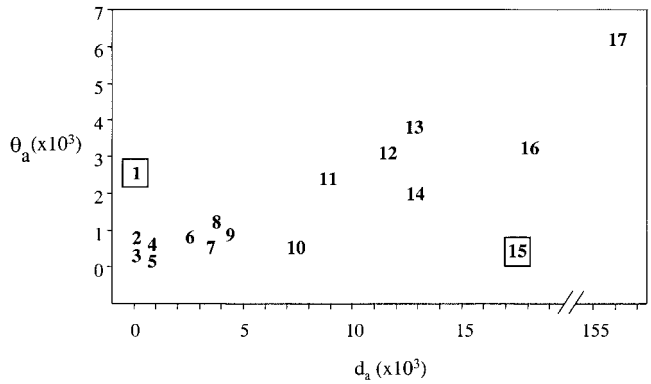

FIGURE 5.—Relationship between amino acid polymorphism (θ_a) and amino acid divergence (d_a) for 17 *D. melanogaster* loci. Plotted loci are the following: (1) *Pgm*, this study; (2) *Sod*, HUDSON *et al.* (1994); (3) *Tpi*, HASSON *et al.* (1998); (4) *Pgi*, JOHN H. MCDONALD, personal communication; (5) *per*, KLIMAN and HEY (1993a); (6) *boss*, AYALA and HARTL (1993); (7) *Adh*, MCDONALD and KREITMAN (1991); (8) *Gld*, HAMBLIN and AQUADRO (1997); (9) *runt*, LABATE *et al.* (1999); (10) *Pgd*, BEGUN and AQUADRO (1994); (11) *Amy-p* and (12) *Amy-d*, INOMATA *et al.* (1995); (13) *Est6*, COOKE and OAKESHOTT (1989); KAROTAM *et al.* (1995); (14) *ase*, HILTON *et al.* (1995); (15) *G6pd*, EANES *et al.* (1993); (16) *Mst26Aa* and (17) *Mst26Ab*, AGUADÉ *et al.* (1992). Boxed loci, *Pgm* (1) and *G6pd* (15), show significant deviations from neutrality by McDonald-Kreitman tests (both $P < 0.001$).

unseen as polymorphisms. A significant and disproportionate number of these fixed differences have accumulated in the *D. simulans* lineage, and this is consistent with a larger effective population size and increased efficacy at fixing adaptive protein polymorphism (EANES *et al.* 1996). By showing no amino acid fixations, yet a significant excess of amino acid polymorphism, the pattern for *Pgm* is unique among *D. melanogaster* loci.

Excessive intraspecific amino acid variation has been reported in several studies involving mtDNA (BALLARD and KREITMAN 1994; NACHMAN *et al.* 1996; RAND and KANN 1996) and three explanations have been proposed for this excess: (1) Replacement mutations are typically slightly deleterious and as a consequence contribute disproportionately to polymorphism and divergence (KIMURA 1983; OHTA 1992). (2) There has been a recent period of relaxed constraint that would affect immediate within-species amino acid variation, but not long-term substitution rates (KENNEDY and NACHMAN 1998). (3) Positive selection maintains amino acid polymorphism by fluctuating over time and space, and these fluctuations either retard or do not impact fixation probabilities (GILLESPIE 1991, 1994).

There is increasing evidence that the classical neutral model (KIMURA 1983) is inadequate to explain certain patterns in DNA evolution (KREITMAN 1996; OHTA 1996). A more realistic relaxed model assumes a spectrum of selection where many segregating mutations possess selection coefficients near the reciprocal of the effective population size. As population size decreases, selection becomes increasingly less effective at countering drift and at the limit behaves as in the classical neutral model. This slightly deleterious or nearly neutral model is difficult to test explicitly because it involves an unknown spectrum of selection coefficients for classes with unknown mutation rates. It can be shown that as selection intensity or population size changes, fixation rates are more responsive than polymorphism levels (KIMURA 1983). Nevertheless, under both models a positive correlation is still expected between levels of polymorphism and divergence. The relationship will be complex in contrasts between silent and replacement sites, because silent and replacement substitutions will have different frequency distributions under selection (AKASHI 1999). Therefore, the classical neutral model may be more applicable to silent site variation, while replacement variation may follow a nearly neutral model where most variation is slightly deleterious. The deviations generated by this mixed model cannot explain apparent deficiencies of amino acid polymorphisms (or excess fixations) as seen for *G6pd* and *Adh* (MCDONALD and KREITMAN 1991; EANES *et al.* 1993). However, excess replacement polymorphisms, as seen in mtDNA genes (BALLARD and KREITMAN 1994; RAND *et al.* 1994; NIELSEN and WEINREICH 1999) and *Pgm* (this study), may be consistent with a nearly neutral interpretation. In the case of mtDNA, this cannot be further addressed due to the lack of suitable comparisons (independent genes with the same N_e), but for *Pgm*, contrasts with other nuclear genes are informative. If a nearly neutral model serves as a global explanation for amino acid variation and *Pgm* is typical, then one must reconcile why 10 of the genes with lower amino acid polymorphism than *Pgm* in Figure 5 all possess higher levels of amino acid divergence. On the other hand, the 5 genes with comparable levels of amino acid polymorphism to *Pgm* show very high levels of amino acid divergence. The bottom line is that *Pgm* appears to possess too much amino acid polymorphism.

The general observation of lower levels of amino acid polymorphism in *D. simulans* is extended to *Pgm* (Moriyama and Powell 1996). Although there are five amino acid polymorphisms in our sample, they are found only with the allozyme alleles and likely segregate as low frequency polymorphisms in larger samples (additional copies of the allozyme alleles show these amino acid polymorphisms are rare, but not singletons). Like the amino acid polymorphisms, many silent polymorphisms at *Pgm* segregate as singletons on rare haplotypes. This odd DNA haplotype structure in *D. simulans* is possibly explained by historical population structure and subsequent admixture (Ballard and Kreitman 1994; Hasson *et al.* 1998; Hamblin and Veuille 1999). The difference in levels of polymorphism between *D. simulans* and *D. melanogaster* has been explained by a reduced effective population size in *D. melanogaster* resulting in higher levels of replacement polymorphism (if the majority of this class are slightly deleterious mutations), but lower levels of synonymous site variation, as seen here. This is further supported by the observation of increased levels of amino acid fixation in many genes in the *D. melanogaster* lineage (Akashi 1995, 1996; Eanes *et al.* 1996). Increased levels of amino acid fixation would be indicative of a long-term reduction in population size, but one must then explain for *Pgm* the absence of amino acid fixation in the face of the abundance of amino acid polymorphism.

Other features of the polymorphism at the *Pgm* locus are inconsistent with a recent decrease in population size and a release of deleterious mutation into the pool of amino acid polymorphisms. Recent contractions in population size with subsequent expansion result in the distortion of the frequency spectrum tending toward mutations at low frequencies, and in the case of mtDNA variation there are typically excesses of replacement polymorphism in the singleton class (Nachman *et al.* 1994, 1996; Rand and Kann 1996; Hasegawa *et al.* 1998; Nielsen and Weinreich 1999), consistent with a slightly deleterious model. However, many of the *Pgm* amino acid replacements in the *D. melanogaster* lineage are found at substantial frequencies (only 6 of the 12 are singletons), and both the Tajima test and the Fu and Li test find no significant skew in either the frequency distribution of amino acid variation or silent site variation. Because metabolic enzymes exhibit some of the highest codon bias for *D. melanogaster* genes (Kliman and Hey 1993b), we were also interested in examining silent site fixations between *D. melanogaster* and *D. simulans* for any significant lineage-specific divergence (Akashi 1995; Eanes *et al.* 1996). A significant excess of fixed unpreferred codons would predict historically relaxed functional constraint at the *Pgm* locus. We find no significant difference in silent site fixation between lineages, nor is there evidence for the fixation of preferred or unpreferred codons at *Pgm* in either lineage when *D. yakuba* is used as an outgroup. Because codon bias for *Pgm* is only slightly above average [codon adaptation index (CAI) = 0.51, relatively low for a metabolic gene], we might not expect a significant trend for the fixation of preferred or unpreferred codons. Weak purifying selection acting on amino acid polymorphisms will not affect the polymorphism frequency spectrum of linked silent site variation (Akashi 1999; Przeworski *et al.* 1999). Therefore, while an analysis of polymorphism frequency distributions for preferred and unpreferred codons is informative for demographic inferences, it cannot explain the excess of amino acid polymorphism.

A recent decrease in population size should also be reflected in other genes since all are demographically influenced in the same fashion. However, there is no trend toward an excess of singletons for other *D. melanogaster* genes (Akashi 1995; Moriyama and Powell 1996). Thus, a period of relaxed functional constraint for PGM would have to demonstrate that this period has been fairly recent, given the absence of amino acid fixation at this locus. Given the recent colonization of North America by *D. melanogaster* (David and Capy 1988), the pattern of variation at *Pgm* may be an artifact of a founding population that has undergone a bottleneck. The Zimbabwe population represents an additional population for comparison to patterns of variation in the North American sample. Compared to cosmopolitan populations, estimates of nucleotide variation for other genes from Zimbabwe tend to show elevated levels of silent polymorphism, which is consistent with a much larger historical ancestral population size in Africa (Begun and Aquadro 1993; Eanes *et al.* 1996). Both Figure 1 and Figure 3 show that these two populations share the same genealogy at this locus. There is neither evidence for geographic clustering due to fixed differences nor an absence of shared polymorphism between population samples. If the analysis of polymorphism at *Pgm* is indicative of a decrease in population size for recently colonized areas, then we might expect lowered levels of variation in our North American sample. However, silent site variation is comparable in the two samples for *Pgm*, and the Zimbabwe population exhibits the same high level of amino acid polymorphism (Table 1). Although both samples harbor some private alleles for both silent and replacement polymorphism, the level of nucleotide diversity between samples is almost identical to the level of variation within samples. While this is generally unusual for comparisons between Zimbabwe and non-Zimbabwe populations (Begun and Aquadro 1993), these patterns of variation are typical of a few *D. melanogaster* loci (*Acp26Aa*, Tsaur *et al.* 1998; *cecropin*, Clark and Wang 1997). This discrepancy suggests selection had historically determined variation at several, but not all, loci in ancestral African populations before *D. melanogaster* colonized cosmopolitan regions (Begun *et al.* 1999). Finally, there is no excess of singletons for either class of mutations in the

North American or the Zimbabwe sample, as would be expected if amino acid polymorphism had recently accumulated at this locus or if there had been a recent bottleneck of silent site variation.

A selection model that manages to maintain amino acid polymorphisms within *D. melanogaster* while limiting amino acid fixation is required to explain the pattern of replacement mutation at the *Pgm* locus. Balancing selection predicts the persistence of lineages in the population longer than expected under genetic drift, and associated lineages should have elevated levels of between-lineage variation because of hitchhiking near the selected polymorphisms (Kaplan *et al.* 1988; Kreitman and Hudson 1991). At *Pgm*, there is no evidence of an excess of silent variation associated with the amino acid polymorphisms. The textbook case for balancing selection has become the study of *Adh* in *D. melanogaster* (Kreitman 1983; Kreitman and Hudson 1991), where a peak of excess silent polymorphism is found around the *F/S* polymorphism. This pattern of variation is novel and may be the exception rather than the rule for *D. melanogaster* loci, even where there is good ancillary evidence for natural selection (Kreitman and Akashi 1995; McDonald 1998). In addition, evidence suggests that this excess of silent site variation arose prior to the *Fast* allele, and the peak of polymorphism around the *F/S* polymorphism is purely coincidental (Begun *et al.* 1999). The difficulty of using silent polymorphism in detecting a pattern of balancing selection for *Pgm* is further complicated by the many amino acid polymorphisms. Because it is also unnecessary to invoke selection to explain all amino acid variation at the *Pgm* locus (some variation may have no adaptive value), it can be difficult to separate neutral background noise from the selected signal. There are also other reasons why an excess of silent site polymorphism may not always accompany a balanced polymorphism. If the local recombination rate is sufficiently high or the selected site is relatively recent, the signature of balancing selection will not be reflected in the genealogical structure of linked neutral variation (Hudson *et al.* 1994; Navarro *et al.* 2000). Balancing selection may also be difficult to detect if much adaptive polymorphism is short lived. In this case, the accumulation of neutral variation will be negligible and an excess of silent polymorphism will not be observed. Instead, adaptive polymorphisms that are short lived and relatively recent may cause linked neutral variation to be skewed toward an excess of rare alleles (Gillespie 1994, 1997).

Models with fluctuating selection might best explain the excess of amino acid polymorphisms seen at the *Pgm* locus in *D. melanogaster*. The long-term turnover of amino acid polymorphisms may be the result of selection operating in response to fluctuating and ephemeral environments. Because adaptive episodes or environments may be short lived, amino acid polymorphisms may be driven to intermediate frequencies by positive selection, but fail to reach fixation as they are replaced by a continuous traffic of amino acid polymorphism. This incomplete fixation process generates overlapping hitchhiking events, where silent variation linked to adaptive amino acid polymorphism is reduced. Wayne *et al.* (1996) explain an excess of amino acid polymorphism at the *ref(2)P* locus in *D. melanogaster* as an adaptive response for resistance to the rapidly evolving sigma virus. However, the level of overall amino acid fixation for this locus is also relatively high. The fact that amino acid fixation is rare for *Pgm* suggests that periods of environmental fluctuation are very short relative to the strength of positive selection acting on amino acid polymorphism.

These episodes may be difficult to detect in DNA sequences because, while hitchhiking under fluctuating selection may decrease heterozygosity, the resulting frequency distributions may not deviate from neutrality (Gillespie 1994, 1997; Braverman *et al.* 1995). Although the Tajima and the Fu and Li tests were not significant for the overall gene analysis, the trend of negative D values suggests an excess of rare alleles at *Pgm*. While these statistical tests lack the power to detect the effect of purifying selection on linked neutral variation (Golding 1997; Neuhauser and Krone 1997; Przeworski *et al.* 1999), they can detect distortions in genealogical structure resulting from recent and recurrent selective sweeps (Braverman *et al.* 1995). We were interested in whether there possibly is an underlying structure to the sequence variation that can explain the apparent excess of rare alleles (Depaulis *et al.* 1999). A subset of *Pgm* alleles with significantly fewer silent polymorphisms than expected under neutrality would be indicative of a mutation recently favored by directional selection. Because amino acid mutations have functional implications, we were especially interested in accessing the potential of specific amino acid polymorphisms to distort the genealogical structure in our sample. From the polymorphism data in Figure 1 and the genealogy structure implied in Figure 3, there are two lineages of *Pgm* sequences in the North American CRS that are defined by the amino acid polymorphism at nucleotide site 2055 (V484L). While our sample contains many amino acid polymorphisms, we were specifically interested in the site 2055 polymorphism because it is the only amino acid polymorphism that is intermediate in frequency and also lies at the base of the gene tree. For this analysis, lineage A is defined by a *G* at nucleotide site 2055 and lineage B is defined by a *T*. Both the Tajima test and the Fu and Li test were used to test the frequency distributions of both silent and replacement polymorphisms independently for these two lineages to detect deviations from that expected under a strictly neutral model (Table 4). Neither test found any deviations from neutrality for lineage B, while both tests found the frequency distributions of both silent and replacement polymorphisms significantly de-

TABLE 4

Summary parameter estimates and neutrality tests for the site 2055 lineages

Sample	n	θ	π	Tajima's D	Fu and Li's D
Lineage A (silent)	14	0.0031	0.0014	−1.80*	−2.61**
Lineage A (rep)	14	0.0015	0.0007	−1.90*	−2.25*
Lineage B (silent)	7	0.0090	0.0077	−0.77	−0.15
Lineage B (rep)	7	0.0016	0.0017	0.29	0.76

Lineage A and B refer to the two alleles designated by the amino acid polymorphism at nucleotide site 2055 (see DISCUSSION). Tajima's D and Fu and Li's D are shown for tests of the silent and replacement polymorphism frequency distributions for all four samples. *$P < 0.05$, **$P < 0.01$.

viated from neutrality for lineage A. Not only is less variation associated with lineage A, but also low frequency variants are found in significant excess for this lineage. It is unclear whether selection is acting upon the amino acid polymorphism at this defining site, or a site within or linked with this haplotype. However, it is clear that this specific *Pgm* amino acid haplotype may have recently increased in frequency. A fundamental feature of the coalescence process is that allele age, frequency, and intra-allele variability are correlated (SLATKIN and RANNALA 1997; WIUF and DONNELLY 1999). It should be emphasized that the pattern of variation associated with this *Pgm* amino acid haplotype is not the case of a new mutation, since the A lineage bears the ancestral state at nucleotide site 2055. An analysis of this variation along a latitudinal cline shows this haplotype markedly increasing in temperate climates (B. C. VERRELLI and W. F. EANES, unpublished data). This pattern of variation is consistent with a model of fluctuating selection where a certain amino acid haplotype may increase in frequency and accumulate little variation over a short period of time, but is limited from reaching fixation.

Association with the *In(3L)P* inversion: Inversion polymorphisms are a pervasive feature of Drosophila genomes and their role in structuring genic variation by suppressing recombination is an enduring question. In this regard the *Pgm* locus occupies a potentially interesting chromosomal position. We had initially mapped the cytological position of *Pgm* by *in situ* hybridization to third chromosome bands 72D1–5, which are immediately inside the proximal breakpoint (73E3) of the *In(3L)P* inversion. With the completion of the entire Drosophila genome sequence, the exact location of *Pgm* is determined to be ~180 kb inside the proximal inversion breakpoint. Despite this close proximity and the apparent old age of the *In(3L)P* inversion (WESLEY and EANES 1994), there are no fixed differences between *Pgm* alleles on different arrangements, and there are also shared polymorphisms indicating genetic exchange. Recombination is expected to be rare near inversion breakpoints because of difficulty in homologous pairing in inversion heterozygotes (KRIMBAS and POWELL 1993), and this restriction has apparently facilitated fixations between the standard and the *In(3L)P* arrangements in the immediate vicinity of the breakpoints (WESLEY and EANES 1994; HASSON and EANES 1996). Gene conversion is the likely mechanism for genetic exchange near inversion breakpoints, and it can explain the exchange of at least two small tracts of *Pgm* sequence between standard and inverted arrangements, and the three *In(3L)P* sequence clusters that are distributed among the standard arrangements in Figures 1 and 3. Apparent gene conversion has been reported for *rp49* on standard and inverted arrangements in *D. subobscura* (ROZAS and AGUADÉ 1994; ROZAS *et al.* 1999). This evidence of exchange, combined with the fact that the same PGM allozyme allele arises from different amino acid changes, explains the lack of association between PGM allozyme variation and this inversion in earlier studies (LANGLEY *et al.* 1974, 1977).

Although gene conversion has contributed to exchange between arrangements, it is possible to infer the ancestral *Pgm* sequence initially captured by the inversion event. Two *Pgm* sequences found on inverted chromosomes DPF95 48.2 and HFL97 93 appear ancestral to all standard sequences in the genealogy in Figure 3. This is in accord with HASSON and EANES (1996), who showed that alleles on the *In(3L)P* chromosomes coalesce earlier in time than all alleles on the standard chromosomes. This basal relationship may be expected if this inversion is under balancing selection (METTLER *et al.* 1977; KNIBB *et al.* 1981). *Pgm* alleles on *In(3L)P* chromosomes do not show elevated levels of polymorphism nor are there any fixed differences between arrangements; however, gene conversion at the *Pgm* locus will erode any effect of balancing selection. Although there is evidence of exchange between arrangements, *Pgm* alleles on inverted chromosomes that cluster ancestral to all other alleles may represent the allele first captured by the inversion.

The pattern of linkage disequilibrium: The linkage disequilibrium analysis of the entire 2354-bp region of *Pgm* shows that many sites are strongly correlated. This chromosomal region is predicted to have moderate rates of recombination (see HUDSON and KAPLAN 1995). Using the estimate of the recombination parameter $C = 4Nc$ from HUDSON (1987), which is based on the vari-

ance in the number of pairwise differences, we obtain a locus-specific value of $C = 11.20$. Dividing this value by our locus-specific estimate of θ, the effective number of recombination events per mutation event (c/μ) is estimated to be 1.37. The estimates for a few *D. melanogaster* loci, *Mlc1* (13.4; Leicht *et al.* 1995) and *Tpi* (8.89; Hasson *et al.* 1998), result in larger values, while *G6pd* (1.7; Eanes *et al.* 1996), *Adh* (1.6; Hudson 1987), and *Sod* (0.80; Hudson *et al.* 1994) are comparable to *Pgm*. There is also direct evidence of recombination in our sample from inspection of Figures 1 and 3, and using the criteria of Hudson and Kaplan (1985) at least five recombination events are detected in the Davis Peach Farm CRS. The absence of amino acid fixation, yet a normal level of silent site divergence at *Pgm*, implies that despite ubiquitous linkage disequilibrium across the *Pgm* locus, recombination has been sufficient that amino acid polymorphism has been historically uncoupled from silent site polymorphism.

It is likely that much of this disequilibrium has a recent mutational origin; an immediate association arises between a new mutation and sites on the allele that the new mutation first appears. As new mutations persist as low frequency variants, only after sufficient time will these initial associations between sites be reduced by recombination. Our sample shows evidence of recombination, but these involve intermediate frequency mutations. Much of the disequilibrium appears due to a single amino acid haplotype (as defined by a *G* at nucleotide site 2055) at high frequency with a significant excess of low frequency variants. Although the entire sample shows a trend toward rare mutations and association between sites, the strong disequilibrium associated with this amino acid haplotype is likely the result of recent directional selection.

Within-species allozyme heterogeneity: This study clearly shows that the PGM allozyme mobility classes can be heterogeneous mixtures of amino acid replacements, where amino acid replacements often converge on the same electrophoretic mobility. This within-allozyme heterogeneity had been predicted from thermostability studies on PGM (Trippa *et al.* 1976, 1978) and may explain the failure to detect functional differences in kinetic studies (Fucci *et al.* 1979; Carfagna *et al.* 1980), the absence of linkage disequilibrium between allozyme mobility classes and the *In(3L)P* inversion (Langley *et al.* 1974, 1977), and the reported absence of latitudinal clines for PGM allozyme alleles (Oakeshott *et al.* 1981). The observation of three allozyme mobility alleles shared between *D. melanogaster* and *D. simulans* warranted study. The *Medium* allozyme allele in *D. simulans* is identical in amino acid sequence to one of the *D. melanogaster Medium* allozyme alleles, so both species still share a common ancestral amino acid sequence from which numerous alleles emerge. However, the other *Medium* alleles and derived *Fast* and *Slow* mobility classes have different mutational origins in the two species. Katz and Harrison (1997) found that although two distantly related cricket species in the genus Gryllus shared six allozyme mobility alleles for phosphoglucose isomerase, there were also no shared amino acid polymorphisms between the two species. Both studies emphasize how tenuous studies of enzyme variation and function can be when based solely on electrophoretic studies, especially in highly polymorphic allozyme loci.

Although *D. yakuba* does not share the same allozyme mobility alleles with *D. melanogaster* and *D. simulans*, it exhibits the same high level of allozyme variability. Three allozyme mobility alleles, two of which are high in frequency, were found in a screen of only 30 isofemale lines from the West African population sample. Our sample of *D. yakuba Pgm* sequences segregates four replacement polymorphisms (which are all responsible for electrophoretic differences), and like *D. melanogaster* and *D. simulans*, different amino acid replacements result in the same allozyme mobility alleles (our unpublished data). A preliminary analysis of this sample also shows a normal level of silent site divergence between the *D. melanogaster-simulans* lineage and the *D. yakuba* lineage at *Pgm*, yet a low number of amino acid fixations.

Adaptive protein evolution in the glycolytic pathway: *D. melanogaster* has been an important model for studying selection on enzyme polymorphisms dating back to the earliest of allozyme studies. While there is evidence for selection on some metabolic enzymes, it is of fundamental importance to understand why specific enzymatic points in the pathway possess protein polymorphisms and rapid evolution, while others do not. Because of their position and intrinsic ability to allocate substrate into competing minor pathways (LaPorte *et al.* 1984; Keightley and Kacser 1987; Keightley 1989), molecular variation in branch point enzymes may be subject to adaptive evolution. The location of PGM at a branch point at the head of the metabolic pathway and its high allozyme polymorphism implicated adaptive protein evolution. Like *G6pd* (another branch point enzyme competing for the same substrate), *Pgm* shows significant patterns of adaptive replacement polymorphism, but unlike *Adh* (Kreitman 1983; Kreitman and Hudson 1991) and possibly *G6pd* (McDonald 1998), *Pgm* does not exhibit the patterns of silent site variation expected of long-term balancing selection. It is possible that selection maintains diversifying variation in the competing storage of glycogen or that amino acid polymorphisms at *Pgm* arise from structure-function tradeoffs between enzyme activity and stability (see Eanes 1999). The absence of any geographic variation for PGM allozymes was an unusual observation for a metabolic enzyme in *D. melanogaster* (Oakeshott *et al.* 1981); however, we find that amino acid haplotypes at the *Pgm* locus in *D. melanogaster* show strong and significant clines with latitude (B. C. Verrelli and W. F. Eanes, unpublished data). This adds compelling evidence for the selective maintenance of amino acid polymorphism at this locus.

The authors thank Ing-Nang Wang and Mike McCartney for their help in the early stages of the project with the recovery of the initial *Pgm* transcript. We also thank John H. McDonald for his interest in this project and useful comments on an earlier version of the manuscript. Jody Hey and two anonymous reviewers provided helpful criticism in revision. This research was supported by National Science Foundation grant DEB9318381 and U.S. Public Health Service grant GM-45247 to W.F.E. This is contribution number 1074 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

LITERATURE CITED

Aguadé, M., N. Miyashita and C. H. Langley, 1992 Polymorphism and divergence in the *Mst26A* male accessory gland gene region in Drosophila. Genetics **132:** 755–770.

Akashi, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics **139:** 1067–1076.

Akashi, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics **144:** 1297–1307.

Akashi, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221–238.

Anderson, P. R., and J. G. Oakeshott, 1984 Parallel geographical patterns of allozyme variation in two sibling Drosophila species. Nature **308:** 729–731.

Argos, P., M. G. Rossmann, U. M. Grau, H. Zuber, G. Frank *et al.*, 1979 Thermal stability and protein structure. Biochemistry **18:** 5698–5703.

Ayala, F. J., and D. L. Hartl, 1993 Molecular drift of the bride-of-sevenless (*boss*) gene in Drosophila. Mol. Biol. Evol. **10:** 1030–1040.

Ballard, J. W. O., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of Drosophila: selection and geographic differentiation. Genetics **136:** 155–171.

Begun, D. J., A. J. Betancourt, C. H. Langley and W. Stephan, 1999 Is the *Fast/Slow* allozyme variation at the *Adh* locus of *Drosophila melanogaster* an ancient balanced polymorphism? Mol. Biol. Evol. **16:** 1816–1819.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

Burkhart, B. D., E. Montgomery, C. H. Langley and R. A. Voelker, 1984 Characterization of allozyme null and low activity alleles from two natural populations of *Drosophila melanogaster*. Genetics **107:** 295–306.

Carfagna, M., L. Fucci, L. Gaudio, G. Pontecorvo and R. Rubino, 1980 Adaptive value of PGM polymorphism in laboratory populations of *Drosophila melanogaster*. Genet. Res. **36:** 265–276.

Chakrabartty, A., J. A. Schellman and R. L. Baldwin, 1991 Large differences in the helix propensities of alanine and glycine. Nature **351:** 586–588.

Clark, A. G., and L. Wang, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713–724.

Cooke, P. H., and J. G. Oakeshott, 1989 Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:** 1426–1430.

Dai, J.-B., Y. Liu, W. J. Ray, Jr. and M. Konno, 1992 The crystal structure of muscle phosphoglucomutase refined at 2.7-angstrom resolution. J. Biol. Chem. **267:** 6322–6337.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

Dawson, D. M., and S. Jaeger, 1970 Heterogeneity of phosphoglucomutase. Biochem. Genet. **4:** 1–9.

Depaulis, F., L. Brazier and M. Veuille, 1999 Selective sweep at the *Drosophila melanogaster* Suppressor of Hairless locus and its association with the *In(2L)t* inversion polymorphism. Genetics **152:** 1017–1024.

Eanes, W. F., 1999 Analysis of selection on enzyme polymorphisms. Annu. Rev. Ecol. Syst. **30:** 301–326.

Eanes, W. F., M. Kirchner and J. Yoon, 1993 Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *D. simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, I.-N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. Genetics **144:** 1027–1041.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Fucci, L., L. Gaudio, R. Rao, A. Spano and M. Carfagna, 1979 Properties of the two electrophoretic variants of phosphoglucomutase in *Drosophila melanogaster*. Biochem. Genet. **17:** 825–836.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.

Gillespie, J. H., 1994 Substitution processes in molecular evolution. III. Deleterious alleles. Genetics **138:** 943–952.

Gillespie, J. H., 1997 Junk ain't what junk does: neutral alleles in a selected context. Gene **205:** 291–299.

Golding, G. B., 1997 The effect of purifying selection on genealogies, pp. 271–285 in *Progress in Population Genetics and Human Evolution*, edited by P. Donnelly and S. Tavare. Springer Verlag, New York.

Hamblin, M. T., and C. F. Aquadro, 1997 Contrasting patterns of nucleotide sequence variation at the glucose dehydrogenase (*Gld*) locus in different populations of *Drosophila melanogaster*. Genetics **145:** 1053–1062.

Hamblin, M. T., and M. Veuille, 1999 Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics **153:** 305–317.

Hasegawa, M., Y. Cao and Z. Yang, 1998 Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol. Biol. Evol. **15:** 1499–1505.

Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144:** 1565–1575.

Hasson, E., I.-N. Wang, L.-W. Zeng, M. Kreitman and W. F. Eanes, 1998 Nucleotide variation in the triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **15:** 756–769.

Hilton, H., R. M. Kliman and J. Hey, 1995 Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48:** 1900–1913.

Hirose, M., E. Sugimoto, R. Sasaki and H. Chiba, 1970 Crystallization and reaction mechanism of yeast phosphoglucomutase. J. Biochem. **68:** 449–457.

Hjorth, J. P., 1970 A phosphoglucomutase locus in *Drosophila melanogaster*. Hereditas **64:** 146–148.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

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.

Hyytia, P., P. Capy, J. R. David and R. S. Singh, 1985 Enzymatic and quantitative variation in European and African populations of *Drosophila simulans*. Heredity **54:** 209–217.

Inomata, N., H. Shibara, E. Okuyama and T. Yamazaki, 1995 Evolutionary relationships and sequence variation of α-amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster*. Genetics **141:** 237–244.

Kaplan, N. L., T. Darden and R. R. Hudson, 1988 The coalescent process in models with selection. Genetics **120:** 819–829.

Karotam, J., A. C. Delves and J. G. Oakeshott, 1995 Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling Drosophila species. Genetica **88:** 11–28.

Katz, L. A., and R. G. Harrison, 1997 Balancing selection on electrophoretic variation of phosphoglucose isomerase in two species of field cricket: *Gryllus veletis* and *G. pennsylvanicus*. Genetics **147:** 609–621.

Keightley, P. D., 1989 Models of quantitative variation of flux in metabolic pathways. Genetics **121:** 869–876.

Keightley, P. D., and H. Kacser, 1987 Dominance, pleiotropy, and metabolic structure. Genetics **117:** 319–329.

Kennedy, P., and M. W. Nachman, 1998 Deleterious mutations at the mitochondrial ND3 gene in South American marsh rats (Holochilus). Genetics **150:** 359–368.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.

Kliman, R. M., and J. Hey, 1993a DNA sequence variation at the period locus within and among species of the *Drosophila melanogaster* complex. Genetics **133:** 375–387.

Kliman, R. M., and J. Hey, 1993b Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

Knibb, W. R., J. G. Oakeshott and J. B. Gibson, 1981 Chromosome inversion polymorphisms in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. Genetics **98:** 833–847.

Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

Kreitman, M., 1996 The neutral theory is dead. Long live the neutral theory. Bioessays **18:** 678–683.

Kreitman, M., and H. Akashi, 1995 Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26:** 403–422.

Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

Krimbas, C. B., and J. R. Powell, 1993 Introduction, pp. 1–52 in *Inversion Polymorphism in Drosophila*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.

Labate, J. A., C. H. Biermann and W. F. Eanes, 1999 Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **16:** 724–731.

Langley, C. H., Y. N. Tobari and K.-I. Kojima, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Genetics **78:** 921–936.

Langley, C. H., K. Ito and R. A. Voelker, 1977 Linkage disequilibrium in natural populations of *Drosophila melanogaster*: seasonal variation. Genetics **86:** 447–454.

Langley, C. H., R. A. Voelker, A. J. Leigh-Brown, S. Ohnishi, B. Dickson *et al.*, 1981 Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. Genetics **99:** 151–156.

LaPorte, D. C., K. Walsh and D. E. Koshland, Jr., 1984 The branch point effect. J. Biol. Chem. **259:** 14068–14075.

Leicht, B. G., S. V. Muse, M. Hanczyc and A. G. Clark, 1995 Constraints on intron evolution in the gene encoding the myosin alkali light chain in Drosophila. Genetics **139:** 299–308.

Liu, Y., W. J. Ray, Jr. and S. Baranidharan, 1997 Structure of rabbit muscle phosphoglucomutase refined at 2.4-angstrom resolution. Acta Cryst. D **53:** 392–405.

McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequences in the ratio of polymorphism to divergence. Mol. Biol. Evol. **15:** 377–384.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Mettler, L. E., R. A. Voelker and T. Mukai, 1977 Inversion clines in populations of *Drosophila melanogaster*. Genetics **87:** 169–176.

Mitton, J. B., 1998 *Selection in Natural Populations*. Oxford University Press, New York.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Nachman, M. W., S. N. Boyer and C. F. Aquadro, 1994 Non-neutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics **155:** 685–698.

Neuhauser, C., and S. K. Krone, 1997 The genealogy of samples in models with selection. Genetics **145:** 519–534.

Nielsen, R., and D. M. Weinreich, 1999 The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. Genetics **153:** 497–506.

Oakeshott, J. G., G. K. Chambers, J. B. Gibson and D. A. Willcocks, 1981 Latitudinal relationships of esterase-6 and phosphoglucomutase gene frequencies in *Drosophila melanogaster*. Heredity **47:** 385–396.

Oakeshott, J. G., S. R. Wilson and P. Parnell, 1985 Selective effects of temperature on some enzyme polymorphisms in laboratory populations of *Drosophila melanogaster*. Heredity **55:** 69–82.

Oakeshott, J. G., S. R. Wilson and W. R. Knibb, 1988 Selection affecting enzyme polymorphisms in enclosed Drosophila populations maintained in a natural environment. Proc. Natl. Acad. Sci. USA **85:** 293–297.

Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

Ohta, T., 1996 The current significance and standing of neutral and nearly neutral theories. Bioessays **18:** 673–677.

Przeworski, M., B. Charlesworth and J. D. Wall, 1999 Genealogies and weak purifying selection. Mol. Biol. Evol. **16:** 246–252.

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.

Ray, W. J., Jr., and G. A. Roscelli, 1964 A kinetic study of the phosphoglucomutase pathway. J. Biol. Chem. **239:** 1228–1236.

Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of Drosophila. Proc. Natl. Acad. Sci. USA **91:** 11517–11521.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Rozas, J., C. Segarra, G. Ribo and M. Aguadé, 1999 Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. Genetics **151:** 189–202.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Singh, R. S., and L. R. Rhomberg, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. Genetics **117:** 255–271.

Slatkin, M., and B. Rannala, 1997 Estimating the age of alleles by use of intraallelic variability. Am. J. Hum. Genet. **60:** 447–458.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Trippa, G., C. Santolamazza and R. Scozarri, 1970 Phosphoglucomutase (PGM) locus in *Drosophila melanogaster*: linkage and population data. Biochem. Genet. **4:** 665–667.

Trippa, G., A. Loverre and A. Catamo, 1976 Thermostability studies for investigating non-electrophoretic polymorphic alleles in *Drosophila melanogaster*. Nature **260:** 42–44.

Trippa, G., G. A. Danieli, R. Costa and R. Scozarri, 1977 A new allele at the PGM locus in *Drosophila melanogaster*. Dros. Inf. Serv. **52:** 74.

Trippa, G., A. Catamo, A. Lombardozzi and R. Cicchetti, 1978 A simple approach for discovering common nonelectrophoretic enzyme variability: a heat denaturation study in *Drosophila melanogaster*. Biochem. Genet. **16:** 299–305.

Tsaur, S. C., C. T. Ting and C.-I. Wu, 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of Drosophila: II. Divergence versus polymorphism. Mol. Biol. Evol. **15:** 1040–1046.

Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wayne, M. L., D. Contamine and M. Kreitman, 1996 Molecular population genetics of *ref(2)P*, a locus which confers viral resistance in Drosophila. Mol. Biol. Evol. **13:** 191–199.

Wesley, C. S., and W. F. Eanes, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91:** 3132–3136.

Whitehouse, D. B., J. Tomkins, J. U. Lovegrove, D. A. Hopkinson and W. O. McMillan, 1998 A phylogenetic approach to the identification of phosphoglucomutase genes. Mol. Biol. Evol. **15:** 456–462.

Winnepenninckx, B., T. Backeljau and R. de Wachter, 1993 Extraction of high molecular weight DNA from molluscs. Trends Genet. **9:** 407.

Wiuf, C., and P. Donnelly, 1999 Conditional genealogies and the age of a neutral mutant. Theor. Popul. Biol. **56:** 183–201.

Communicating editor: J. Hey