Adaptive Evolution of Metabolic Pathways in *Drosophila*

J. M. Flowers, E. Sezgin,[1] S. Kumagai, D. D. Duvernell,[2] L. M. Matzkin,[3] P. S. Schmidt,[4] and W. F. Eanes

Department of Ecology and Evolution, Stony Brook University

The adaptive significance of enzyme variation has been of central interest in population genetics. Yet, how natural selection operates on enzymes in the larger context of biochemical pathways has not been broadly explored. A basic expectation is that natural selection on metabolic phenotypes will target enzymes that control metabolic flux, but how adaptive variation is distributed among enzymes in metabolic networks is poorly understood. Here, we use population genetic methods to identify enzymes responding to adaptive selection in the pathways of central metabolism in *Drosophila melanogaster* and *Drosophila simulans*. We report polymorphism and divergence data for 17 genes that encode enzymes of 5 metabolic pathways that converge at glucose-6-phosphate (G6P). Deviations from neutral expectations were observed at five loci. Of the 10 genes that encode the enzymes of glycolysis, only aldolase (*Ald*) deviated from neutrality. The other 4 genes that were inconsistent with neutral evolution (glucose-6-phosphate dehydrogenase [*G6pd*]), phosphoglucomutase [*Pgm*], trehalose-6-phosphate synthetase [*Tps1*], and glucose-6phosphatase [*G6pase*] encode G6P branch point enzymes that catalyze reactions at the entry point to the pentose-phosphate, glycogenic, trehalose synthesis, and gluconeogenic pathways. We reconcile these results with population genetics theory and existing arguments on metabolic regulation and propose that the incidence of adaptive selection in this system is related to the distribution of flux control. The data suggest that adaptive evolution of G6P branch point enzymes may have special significance in metabolic adaptation.

Introduction

The adaptive significance of enzyme variation has been a central topic of interest in population genetics since the earliest surveys of allozyme polymorphism (Lewontin and Hubby 1966; Lewontin 1974). It is now accepted that enzyme variation may frequently be adaptive (Gillespie 1991; Mitton 1997; Eanes 1999) and that metabolic pathways adapt via amino acid substitution and changes in *cis*- or *trans*-acting factors that modify enzyme activity (Crawford and Powers 1989; Pierce and Crawford 1997). Although the importance of protein evolution for metabolic adaptation is not in question, a comprehensive understanding of how natural selection operates in the context of metabolic pathways remains unaddressed (Eanes 1999; Rausher et al. 1999; Cork and Purugganan 2004). There is a growing need to identify what enzymatic steps have been subject to adaptive evolution and to develop a pathway, or network, perspective of how natural selection shapes the evolution of complex metabolic systems.

An important step toward developing a systems-level view is to identify features of metabolic networks that influence the evolution of individual genes. Large-scale studies have reported correlations between network properties and rates of protein evolution. In yeast, highly connected enzymes have a slower rate of amino acid substitution (Vitkup et al. 2006). This correlation suggests that selective constraints on molecular evolution depend on a system property and that the strength of "purifying" selection is contingent on how connected an enzyme is to other enzymes in the network. Unfortunately, the comparative methods frequently used to estimate evolutionary rates in biological networks can only rarely detect "adaptive" protein evolution (Kreitman and Akashi 1995; Yang and Bielawski 2000). As a consequence, very little is known about how adaptive variation is distributed among enzymes in metabolic networks or what structural and/or functional properties of these systems cause some enzymes to evolve adaptively, whereas others are subject to evolutionary constraints. This gap in current understanding of metabolic pathway evolution can be addressed with population-based approaches that contrast nucleotide polymorphism within species with interspecific divergence.

In biochemical pathways, natural selection is expected to target enzymes that control flux (Hartl et al. 1985; Eanes 1999; Watt and Dean 2000). Consequently, where adaptive selection operates in a pathway will depend on the distribution of flux control among steps. Metabolic control analysis proposes that the regulation of flux through a linear biochemical pathway is distributed among enzymes throughout the pathway (Kacser and Burns 1981). In branching pathways, flux allocation may be controlled primarily by enzymes at the intersections of competing pathways (LaPorte et al. 1984; Stephanopoulos and Vallino 1991). Therefore, branch point enzymes may disproportionately regulate, or control, the expression of biochemical phenotypes, such as the energy pools of glycogen or triglycerides that are important for adaptation in natural populations (Eanes 1999).

Mutations that alter enzyme activity will have larger fitness consequences in enzymes with greater flux control (Hartl et al. 1985). As a consequence, natural selection can redistribute flux most effectively by altering the activity of enzymes at branch points. This hypothesis predicts that networks evolve primarily through adaptive changes in enzymes at major pathway intersections and less frequently through mutational changes in remote enzymes.

Addressing hypotheses about adaptive evolution in metabolic networks requires that the enzymatic steps targeted by natural selection be identified. Previously, we used nucleotide polymorphism data from enzyme-coding loci to

[1] Present address: Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland
[2] Present address: Department of Biological Sciences, Southern Illinois University, Edwardsville, Illinois
[3] Present address: Department of Ecology and Evolutionary Biology, University of Arizona
[4] Present address: Department of Biology, University of Pennsylvania

Key words: positive selection, network evolution, population genomics, adaptation, systems biology.

E-mail: jmflower@life.bio.sunysb.edu.

test for departures from neutrality at enzymatic steps in the central metabolic corridor of Drosophila. This work focused on branch point enzymes that metabolize glucose-6-phosphate (G6P). The first evidence for significant deviations from neutrality came from surveys of nucleotide variation at *G6pd* (i.e., *Zw*; Eanes et al. 1993, 1996) and *Pgm* (Verrelli and Eanes 2000). In contrast, patterns of sequence variation at loci that encode the branch point enzymes, *HEX-A* and *HEX-C* (Duvernell and Eanes 2000) and the downstream glycolytic enzyme *TPI* (Hasson et al. 1998) were consistent with neutrality. These studies suggested that adaptive selection may be prevalent, but establishing the general significance of adaptive evolution of branch point enzymes requires an expanded study of enzymes in the pathways that converge at G6P.

In this report, we significantly expand our survey of nucleotide variation to include 17 enzymes of central metabolism. We collected sequence polymorphism and divergence data from *Drosophila melanogaster* and *Drosophila simulans* for 10 additional enzyme-coding loci and assembled all available nucleotide polymorphism data for enzymes of the 5 pathways that intersect at G6P (i.e., glycolytic, gluconeogenic, glycogenic, trehalose, and pentose shunt). We report 3 additional genes that deviate from neutrality in *D. simulans* and now find that 4 of 7 enzymes at the intersection of these pathways are inconsistent with the neutral model in one or both species. This pattern suggests that Darwinian selection on amino acid variation has been an important mechanism of metabolic network evolution in *Drosophila*. Furthermore, the added evidence of adaptive selection at the G6P branch point suggests that evolutionary change at this major pathway intersection may be a means by which the *Drosophila* metabolic network has adapted in response to natural selection.

Materials and Methods
Line Origins and DNA Sequencing

Drosophila melanogaster sequences new to this study were collected from flies homozygous for extracted 2nd, 3rd, or X chromosomes derived from North American populations (i.e., some combination of alleles from New York, Vermont, Florida, Massachusetts, South Carolina, and Virginia; see Duvernell and Eanes 2000; Verrelli and Eanes 2001a). For *Pgk*, *UGP*, *Gapdh2*, and *Eno*, DNA sequences were also obtained from alleles sampled from Zimbabwe (Kreitman M, personal communication). *Drosophila simulans* sequences were obtained from a set of inbred lines collected from locations in the eastern US. Collection data for these lines have been reported previously (Verrelli and Eanes 2001a). A single *Drosophila yakuba* sequence was obtained either from an inbred line of unknown origin or from the whole genome sequence (http://genome.wustl. edu/). All sequences were obtained by polymerase chain reaction amplification of genomic DNA followed by direct sequencing. The numbers of alleles sequenced for each gene are reported in table 1. All sequences consist of the full-length cds, with the possible exception of small exons at the extreme 5′ end of some genes.

DNA sequences were collected for glyceraldehyde-3-phosphate dehydrogenase paralogs (*Gapdh1*, *Gapdh2*), aldolase (*Ald*), phosphoglyceromutase (*Pglym78*), phosphogluconate kinase (*Pgk*), enolase (*Eno*), UDP-glucose pyrophosphorylase (*UGP*), trehalase (*Treh*), trehalose-6-phosphate synthetase (*Tps1*), and *CG15400* (hereafter, *G6pase*). This gene has been identified as the putative ortholog to the human gene (G6PC) encoding glucose-6-phosphatase (Hubbard et al. 2006) and is predicted to have glucose-6-phosphatase activity (Grumbling et al. 2006). Sequences from these genes are new to this study and have been deposited in GenBank under accession numbers DQ863952–DQ864255.

Data for triosephosphate isomerase (*Tpi*; Hasson et al. 1998), the muscle- and fat body-specific hexokinases (*Hex-A* and *Hex-C*; Duvernell and Eanes 2000), *G6pd* (Eanes et al. 1993, 1996), phosphoglucomutase (*Pgm*; Verrelli and Eanes 2000), and 6-phosphogluconate dehydrogenase (*Pgd*; Begun and Aquadro 1994; Begun and Whitley 2000) have been reported previously. GenBank accession numbers for unpublished phosphoglucoisomerase (*Pgi*) sequences are L27539–L27555 and U20556–U20575. Central metabolic enzymes that are testes specific in their expression (Grumbling et al. 2006) were not included in the analysis.

DNA Sequence Analysis

McDonald–Kreitman (MK) tests were conducted in DnaSP v. 4.10 (Rozas et al. 2003). Fixations were inferred from comparison of *D. melanogaster* and *D. simulans* sequences only. Fixations occurring along the branch to *D. yakuba* were excluded to minimize the effects of multiple hits and to reduce the effect of mutations segregating within *yakuba* on divergence estimates. In lineage-specific MK tests, fixations were assigned to the *melanogaster* or *simulans* lineages by parsimony with a script written in Perl by J.M.F. and S.K. All MK tests were conducted on combined African and North American alleles, where applicable. Ancestral codons were reconstructed base-by-base using *D. yakuba* as an outgroup. The parsimony criterion utilized all mutations in the 3 species alignment (i.e., mutations segregating in *melanogaster* and *simulans* were considered when inferring the most parsimonious ancestral states). Codons with multiple substitutions in the same lineage were treated using the method of Pamilo and Bianchi (1993)—when equal number of mutations defined alternative pathways, the pathway with fewer nonsynonymous changes was accepted. Preferred and unpreferred codons were defined according to Sharp and Lloyd (1993) and polarized by the same parsimony criterion. Statistical significance of 2×2 contingency tables was assessed with 2-tailed Fisher's exact tests. Rates of nucleotide substitution were estimated from pairwise alignments of *D. melanogaster* and *D. simulans* sequences. Alignments were generated with MUSCLE (Edgar 2004) using sequences from the whole genome (Grumbling et al. 2006) or by haphazardly selecting an allele from our polymorphism data sets. Rates were estimated by the approximate method of Comeron (1995) using K-estimator v. 6.1 (Comeron 1999). This

Table 1
McDonald–Kreitman Tests of Neutrality

Gene	N[a]		Polymorphic		Fixed		P[b]
	melanogaster	*simulans*	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous	
G6P branch point							
Tps1	18	9	54	1	47	15	0.0003[c]
Hex-A	18	14	8	0	17	0	1.000
Hex-C	50	14	26	4	27	2	0.671
Pgm	43	13	82	21	33	0	0.0021[c]
G6pase	23	8	26	3	22	15	0.011
G6pd	50	12	52	3	27	22	0.000002[c]
Pgi[d]	21	14	34	5	24	1	0.391
Nonbranch point							
UGP	26	10	46	3	21	0	0.549
Treh	21	9	40	14	19	16	0.068
Pgd	1	7	15	2	19	2	1.000
Ald	12	12	18	1	11	6	0.037
Tpi	25	9	39	3	5	0	1.000
Gapdh1	12	10	6	2	9	1	0.559
Gapdh2	16	10	2	1	12	0	0.200
Pgk	25	12	58	3	9	2	0.165
Pglym78	16	12	21	4	8	0	0.550
Eno	24	13	38	1	14	3	0.079

[a] Number of alleles sequenced.
[b] Significance determined with a 2-tailed Fisher's exact test.
[c] Significant after sequential Bonferroni correction.
[d] Sample size represents the number of genomes sequenced.

allowed for comparison with substitution rates reported for 419 genes by Shapiro et al. (2007).

Entries from table 1 were used to estimate the scaled selection parameter, $2N_e s$, by the Poisson random field method of Bustamante et al. (2002). The analysis was conducted using mkprf and implemented by the Cornell Computational Biology Service Unit at http://cbsuapps.tc.cornell.edu/mkprf.aspx. Markov Chain Monte Carlo sampling of the posterior distribution was performed by retaining 5,000 draws from each of 10 chains. The initial burn-in was 1,000 steps. The mkprf analysis was run without specifying a class structure by using the fixed variance option. *Pgd* was excluded from this analysis because polymorphism data were only available for one species.

Results and Discussion

Under the neutral model of molecular evolution (Kimura 1983), the ratio of polymorphism within and divergence between species should be equivalent for nonsynonymous and synonymous mutations (McDonald and Kreitman 1991). In the absence of selection on synonymous sites, excesses of replacement polymorphism or divergence are consistent with adaptive amino acid changes. The MK tests indicated that *G6pase*, *G6pd*, *Tps1*, and *Ald* individually deviate from neutrality in the direction usually attributed to an excess of replacement fixations, whereas *Pgm* showed a pattern frequently interpreted as an excess of replacement polymorphism (table 1). When fixations are assigned to the *melanogaster* or *simulans* lineages (table 2), the departures from neutrality are consistent with balancing selection in *D. melanogaster* at *Pgm* (Verrelli and Eanes 2000, see below), positive selection at *Tps1*, *G6pase*, and *Ald* in *D. simulans*, and positive selection in both lineages at *G6pd* (Eanes et al. 1993, 1996). When considered

with respect to pathway position, 4 of the 5 genes that deviate from neutrality (*Tps1*, *G6pase*, *G6pd*, and *Pgm*) code for enzymes at the G6P intersection (fig. 1). Neutrality could not be rejected for any of the remaining 12 genes in lineage-specific tests (table 2) or when polymorphisms and fixations were combined across lineages (table 1). After sequential Bonferroni correction, *G6pd*, *Pgm*, and *Tps1* remain significant (table 1).

Analysis of the average strength of selection on amino acid replacement mutations supports our assertion that natural selection has operated on *Pgm*, *G6pd*, *Tps1*, *G6pase*, and *Ald*. The Bayesian method we employed to determine the strength of selection uses the joint nucleotide configurations of the MK table to estimate the posterior probability distribution of the scaled selection coefficient, $2N_e s$ (Bustamante et al. 2002). Estimates of positive coefficients are the highest for *Tps1*, *G6pd*, *G6pase*, and *Ald* and the 97.5 quantiles of the posterior probability distribution of $2N_e s$ for these loci are greater than 0 (fig. 2). This suggests that amino acid substitutions at these loci have been driven to fixation by positive selection. At the opposite extreme, *Pgm* shows the most negative value. Because the Poisson random field model does not provide for balancing selection (Sawyer and Hartl 1992), the interpretation of a negative selection coefficient for *Pgm* is unclear. In *D. melanogaster*, there are 16 replacement polymorphisms and no amino acid fixations since the divergence with *D. simulans* (Verrelli and Eanes 2000, table 2). This excess of replacement polymorphism at *Pgm* is consistent with both neutral and selective processes including a recent relaxation of functional constraints, segregation of slightly deleterious mutations, or balancing selection. The clinal distribution of major amino acid haplotypes in the eastern United States (Verrelli and Eanes 2001a) and the existence of many intermediate frequency amino acid polymorphisms

Table 2
McDonald–Kreitman Tests for Lineage-Specific Deviations from Neutrality

	melanogaster Lineage					*simulans* Lineage				
	Polymorphic		Fixed[a]			Polymorphic		Fixed[a]		
Gene	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous	P[b]	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous	P[b]
G6P branch point										
Tps1	33	0	18	1	0.365	21	1	24	12	0.011
Hex-A	8	0	7	0	1.000	0	0	9	0	1.000
Hex-C	17	4	15	0	0.125	10	0	8	2	0.474
Pgm	19	16	21	0	0.0001	64	5	10	0	1.000
G6pase	15	3	13	8	0.171	12	0	6	4	0.029
G6pd	39	3	11	5	0.030	13	0	12	16	0.0004
Pgi	8	3	16	1	0.269	26	2	7	0	1.000
Nonbranch point										
UGP	28	0	12	0	1.000	20	3	8	0	0.550
Treh	26	2	9	4	0.069	15	12	10	11	0.771
Pgd[c]	3	1	45	7	0.470	15	2	6	0	1.000
Ald	8	0	9	2	0.485	10	1	1	5	0.005
Tpi	28	2	2	0	1.000	12	1	3	0	1.000
Gapdh1	6	2	5	1	1.000	0	0	3	0	1.000
Gapdh2	1	1	10	0	0.167	1	0	2	0	1.000
Pgk	24	1	6	1	0.395	36	2	2	1	0.209
Pglym78	14	4	4	0	0.554	7	1	4	0	1.000
Eno	27	1	6	0	1.000	11	0	7	3	0.090

[a] Fixations have been assigned to *Drosophila melanogaster* and *Drosophila simulans* lineages by parsimony using *Drosophila yakuba* as an outgroup.

[b] Significance determined with a 2-tailed Fisher's exact test.

[c] Values for *Pgd* in *melanogaster* include total fixations between *melanogaster* and *simulans* (Begun and Aquadro 1994).

in *D. melanogaster* (Verrelli and Eanes 2000) suggest the action of balancing selection. However, the level of silent polymorphism is not high for *D. melanogaster* (Moriyama and Powell 1996) nor is no detectable skew in the frequency

spectrum of synonymous mutations, as might be expected for sites flanking an ancient balanced polymorphism (Verrelli and Eanes 2000).

Because most metabolic genes in *Drosophila* show exceptionally high levels of codon bias (Duvernell and Eanes 2000), a possible alternative explanation for the significant MK tests at *G6pd*, *Ald*, *Tps1*, and *G6pase* in *D. simulans* is weak selection on synonymous sites. Evidence for natural selection on synonymous mutations has been reported for *D. simulans* (Akashi 1995; Begun 2001) and is also apparent in our data set. When the genes are divided into higher and lower bias groups by the median effective number of codons (ENC) value, the ratio of polymorphism to
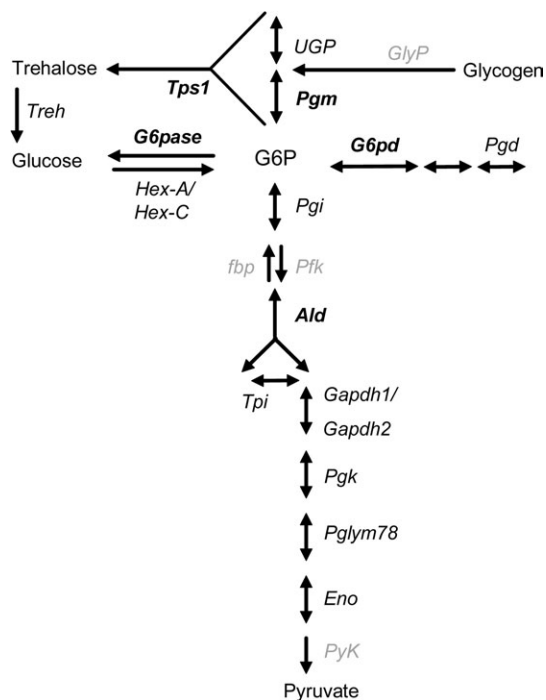

Fig. 1.——The G6P branch point and adjacent pathways in *Drosophila*. Enzymes in boldface deviate from neutrality in individual MK tests. Enzymes in gray were not included in our survey of nucleotide polymorphism.
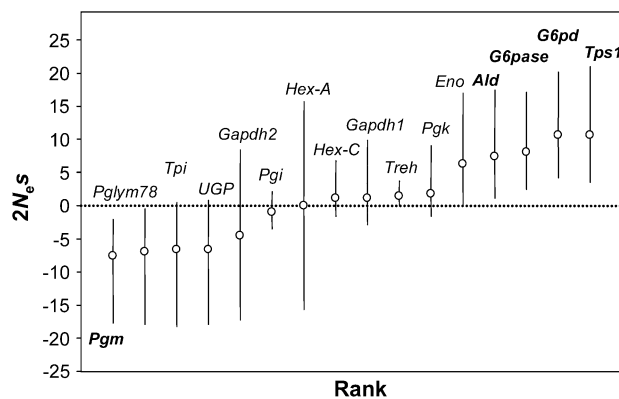

Fig. 2.——Estimates of $2N_es$ for central metabolic genes. Individual points represent the mean of the posterior distribution of $2N_es$. Error bars are 2.5% and 97.5% quantiles of the posterior distribution. Genes in boldface are significant in individual MK tests (table 1). Polymorphism and divergence data used to estimate $2N_es$ are based upon polymorphism data from *melanogaster* and *simulans* in table 1.

Table 3
Polymorphism and Divergence of Polarized Preferred (P) and Unpreferred (U) Mutations in *Drosophila simulans*

Gene	ENC[a]	Polymorphic U → P	Polymorphic P → U	Fixed U → P	Fixed P → U	
Higher bias						
Eno	28.529	1	9	3	2	
Pglym78	31.141	0	4	0	3	
Gapdh1	31.336	0	0	2	0	
Ald	31.381	1	6	0	1	
G6pd	33.349	4	4	4	4	
Tpi	35.037	0	10	1	1	
Gapdh2	35.567	0	1	0	2	
Pgk	36.003	1	25	1	0	
Total	—	7	59	11	13	$P < 0.001$[b]
Lower bias						
Pgd	36.387	3	7	1	5	
Pgi	36.715	7	18	5	1	
Tps1	42.128	0	13	4	14	
Pgm	42.877	14	35	2	5	
Hex-A	42.982	0	0	2	3	
UGP	43.644	4	12	3	4	
Treh	46.57	4	6	4	4	
Hex-C	48.425	1	4	1	5	
G6pase	50.321	2	5	1	2	
Total	—	35	100	23	43	$P = 0.246$[b]

NOTE.—Genes are listed in order of decreasing codon bias. Mutations that could not be polarized by parsimony and those that did not change the preferred or unpreferred status of a codon (i.e., P → P or U → U, Sharp and Lloyd 1993) were excluded from the analysis.

[a] ENC (Wright 1990) were calculated from *Drosophila melanogaster* genome release 4.0 (Grumbling et al. 2006).

[b] Two-tailed Fisher's exact test.

Table 4
Nucleotide Substitution Rates of Metabolic Genes for the *Drosophila melanogaster–Drosophila simulans* Comparison

Gene	K_a		K_s	
Treh	0.0143	(0.0075, 0.0209)	0.0887	(0.0586, 0.1212)
UGP	0.0000	(0, 0)	0.1007	(0.0692, 0.1368)
Tps1	0.0095	(0.0048, 0.0155)	0.1155	(0.0868, 0.1448)
GlyP	0.0014	(0, 0.0029)	0.0605	(0.0419, 0.0806)
G6pase	0.0212	(0.0089, 0.0376)	0.1689	(0.0981, 0.2361)
Pgm	0.0018	(0, 0.0046)	0.1246	(0.0856, 0.1648)
G6pd	0.0206	(0.0125, 0.031)	0.0959	(0.061, 0.1306)
Hex-A	0.0000	(0, 0)	0.0542	(0.0302, 0.0807)
Hex-C	0.0020	(0, 0.0033)	0.1120	(0.0741, 0.1553)
Pgi	0.0009	(0, 0.0027)	0.0812	(0.0538, 0.1147)
Pfk[a]	0.0012	(0, 0.0032)	0.1040	(0.0771, 0.1327)
Ald	0.0034	(0, 0.0076)	0.0392	(0.0152, 0.0662)
Tpi	0.0017	(0, 0.006)	0.0674	(0.028, 0.1177)
Gapdh1	0.0028	(0, 0.0075)	0.0499	(0.0237, 0.0846)
Gapdh2	0.0000	(0, 0)	0.0470	(0.02, 0.0798)
Pgk	0.0022	(0, 0.0036)	0.0706	(0.0416, 0.1012)
Pglym78	0.0000	(0, 0)	0.0829	(0.0406, 0.1272)
Eno	0.0033	(0, 0.0078)	0.0697	(0.0403, 0.1027)
PyK	0.0018	(0, 0.0049)	0.0705	(0.0443, 0.1032)

NOTE.—Rate estimates are Kimura 2-parameter corrected. Values in parentheses are 95% confidence intervals obtained by Monte Carlo simulation (Comeron 1999).

[a] *Drosophila simulans* sequence unavailable. Rate estimates are from the *Drosophila melanogaster–Drosophila sechellia* comparison.

divergence for the higher bias set is 4.54 for total unpreferred and 0.64 for total preferred mutations (table 3). This is significant in a 2×2 contingency test ($P < 0.001$, 2-tailed Fisher's exact test). Conversely, the ratios of polymorphism to divergence for total unpreferred and preferred classes for the lower bias genes are 2.33 and 1.52 ($P = 0.246$), respectively. This rejection of the neutral expectation for the higher bias, but not the lower bias, set is robust to different partitions of the data set. Such deviations from neutrality at synonymous sites has been attributed to a deficit of unpreferred fixations in high bias genes and interpreted as evidence for weak selection on silent mutations (Akashi 1995; Begun 2001).

Consequently, a deficit of synonymous fixations due to selection on silent sites could provide an alternative explanation for what we interpret to be an excess of replacement fixations at *Ald*, *G6pd*, *Tps1*, and *G6pase* in *D. simulans*. This explanation seems unlikely for 3 of the genes. First, *G6pase* shows the lowest codon bias (ENC = 50.32, table 3; the median for *D. melanogaster* is 49.57, Hey and Kliman 2002), the highest synonymous divergence (K_s) in our data set (0.1689; table 4), and has a K_s greater than one standard deviation above the mean for the *melanogaster–simulans* comparison (0.110 ± 0.045; Shapiro et al. 2007). *Tps1* has an intermediate level of codon bias (ENC = 42.13), and a K_s of 0.115, which is third highest in our data set (table 4). This is inconsistent with the hypothesis that negative selection on silent variation explains the pattern at *G6pase* and *Tps1*. Second, under the

mutation-selection-drift model, preferred synonymous mutations are beneficial. Therefore, a conservative approach to testing for positive selection is to consider only preferred mutations in the MK test (Akashi 1995). When preferred silent (table 3) and amino acid replacement (table 2) mutations are considered in an MK test for *G6pd*, there again is a significant excess of amino acid fixations in this lineage ($P < 0.01$, 2-tailed Fisher's exact test). This strongly supports our original assertion that *G6pd* has been subjected to positive selection (Eanes et al. 1993, 1996). Individual tests of *Ald*, *Tps1*, and *G6pase* have few preferred mutations and therefore have weak statistical power to reject neutrality in a comparable test. Third, interpretation of the MK test for *Ald* is problematic because this gene shows extremely high codon bias (uppermost 1% in *D. melanogaster*, Hey and Kliman 2002) and a rate of synonymous divergence that is in the lower 5% of values reported for *melanogaster–simulans* (table 4; Shapiro et al. 2007). This could indicate that synonymous mutations contribute more to polymorphism than divergence at this locus and that selection against synonymous fixations may have distorted the ratio sufficiently to bias the MK test. Although we cannot exclude the possibility that weak selection has contributed to the pattern of polymorphism and divergence at *Ald*, we find it unlikely that selection on codon usage provides a general explanation for the significant MK tests in *D. simulans* (table 2). This reflects a general problem for inferring positive selection in metabolic and other high bias genes in species where there is selection on codon usage (Akashi 1995).

It is conceivable that deviations from neutrality reflect the fixation of slightly deleterious replacement mutations. If *D. simulans* historically had a small population size that recently expanded, the excess of amino acid divergence could

be attributed to the fixation of slightly deleterious variants in the smaller ancestral population (McDonald and Kreitman 1991; Eyre-Walker 2002). A small population size followed by a recent demographic expansion is not supported by population genetic data for *D. simulans*. Most notably, there is an excess of unpreferred silent polymorphism, but not divergence, at synonymous sites (Akashi 1995; Begun 2001, table 3). This suggests that fixations of slightly deleterious amino acid mutations in a small ancestral population are unlikely to explain the pattern in the MK tests. Furthermore, evidence for weak selection on silent sites in *D. simulans* (Akashi 1995; Begun 2001, table 3) greatly restricts the conditions that can lead to an overestimate of adaptive divergence (Eyre-Walker 2002).

Deviations from neutral expectations at loci that encode G6P branch point enzymes suggest that adaptive evolution at this intersection may be important for metabolic adaptation in *Drosophila*. A possible explanation for adaptive selection on branch point enzymes is the localization of flux control at this junction. G6P is an allosteric effector that lies at a common entry point to pathways functioning in the storage, mobilization, transport, and breakdown of carbohydrate. In *D. melanogaster*, there is evidence that G6P branch point enzymes are capable of controlling flux allocation. Flux apportionment to the pentose shunt is extremely sensitive to allozyme variation at *G6pd* (Labate and Eanes 1992), and glycogen storage levels have been reported to depend on *Pgm* genotype (Verrelli and Eanes 2001b). In addition, overexpression of *Tps1* increases trehalose content (Chen et al. 2001), suggesting that this enzyme may control flux in the trehalose pathway.

Mutations that modify activity of controlling enzymes are expected to be visible to natural selection, whereas comparable changes in activity at steps with little or no control are neutral (Hartl et al. 1985). Because branch point enzymes PGM, TPS1, and G6PD exert some measure of control over flux allocation, activity-altering mutations in these enzymes can affect pathway fluxes, modify the phenotype, and are therefore subject to natural selection. Significant MK tests for *Pgm*, *Tps1*, and *G6pd* (Eanes et al. 1993, 1996; Verrelli and Eanes 2000, table 1), latitudinal clines at *G6pd* and *Pgm* in *D. melanogaster* (Oakeshott et al. 1983; Verrelli and Eanes 2001a), and evidence that functional variation at these loci influences biochemical phenotypes supports this expectation. Although the distribution of control among enzymes in these pathways is not well understood, we propose that adaptive substitution at the G6P intersection reflects a greater capability of branch point enzymes to control flux allocation.

Several processes may drive gene-specific adaptive responses in metabolic pathways. First, environmentally driven changes in flux demands and resource partitioning may shift the distribution of control among steps or branches. As the level of control exercised by an enzyme diminishes, alleles with reduced activity become effectively neutral and have an increased probability of fixation (Hartl et al. 1985). These accumulated nearly neutral changes can become deleterious should control return, but can be compensated by subsequent adaptive substitutions that restore activity. Therefore, excesses of allele fixation may reflect a continuous process of network adaptation in which con-

trol shifts and pathway fluxes are fine-tuned to match metabolic demands. Enzymes that lack metabolic control will not experience this process. Alternatively, if control remains static over time, adaptive substitution may occur in a pathway as a consequence of fluctuating population size and adaptive compensatory evolution (Hartl and Taubes 1996). When population sizes contract, slightly deleterious alleles with reduced activity can fix at controlling steps because they are effectively neutral. Subsequent increases in population size promote the fixation of adaptive compensatory changes that restore activity. These compensatory cycles are not expected for enzymes possessing limited control.

In this study, we have reported that 5 of 17 genes surveyed in the central metabolic corridor deviated from neutrality in MK tests, and that 1 enzyme in each of the 5 intersecting pathways showed evidence of natural selection. In addition, our observation that 4 of the 5 genes for which neutrality could be rejected encode G6P branch point enzymes could indicate that natural selection at this major pathway intersection has special significance in metabolic adaptation. When the incidence of selection is compared between G6P branch point and nonbranch point enzymes, 4 of 7 branch point enzymes deviate from neutrality, whereas only 1 of 10 enzymes remote to this intersection show evidence of selection. This is not significant in a 2×2 test ($P > 0.05$, 2-tailed Fisher's exact test), but is marginally significant when departures from neutrality are inferred from Bonferroni-corrected MK tests (i.e., 3 of 7 and 0 of 10 branch point and nonbranch point enzymes reject neutrality; $P = 0.051$). This could suggest that adaptive substitutions are concentrated at the G6P branch point, but additional sampling of enzyme-coding loci will be required to address this question. A recent survey of nucleotide polymorphism in *D. melanogaster* included a 1.1 Kb fragment of *PyK* (Shapiro et al. 2007), which encodes the enzyme pyruvate kinase at the end of the glycolytic pathway. Shapiro et al. (2007) reported no replacement polymorphisms segregating in *D. melanogaster* and no amino acid fixations since the divergence with *simulans* suggesting that this enzyme may also be evolving neutrally. The emerging pattern from these observations and our previous work showing geographic clines in G6P branch point enzymes (Sezgin et al. 2004) is intriguing and implicates enzymes at this intersection in the process of metabolic adaptation. This warrants an expanded survey of nucleotide polymorphism from genes in other pathways and in other species to evaluate the significance of adaptive evolution at the G6P branch point and other principal nodes in metabolic networks.

Literature Cited

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

Begun DJ. 2001. The frequency distribution of nucleotide variation in *Drosophila simulans*. Mol Biol Evol. 18:1343–1352.

Begun DJ, Aquadro CF. 1994. Evolutionary inferences from DNA variation in the 6-Phosphogluconate dehydrogenase locus in natural populations of Drosophila: selection and geographic differentiation. Genetics. 136:155–171.

Begun DJ, Whitley P. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc Natl Acad Sci USA. 97:5960–5965.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. 2002. The cost of inbreeding in Arabidopsis. Nature. 416:531–534.

Chen Q, Ma E, Behar KL, Xu T, Haddad G. 2001. Role of trehalose phosphate synthase in anoxia tolerance and development in *Drosophila melanogaster*. J Biol Chem. 277:3274–3279.

Comeron JM. 1995. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol. 41:1152–1159.

Comeron JM. 1999. K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics. 15:763–764.

Cork JM, Purugganan MD. 2004. The evolution of molecular genetic pathways and networks. Bioessays. 5:479–484.

Crawford DL, Powers DA. 1989. Molecular basis of evolutionary adaptation at the lactate dehydrogenase B locus in the fish *Fundulus heteroclitus*. Proc Natl Acad Sci USA. 86:9365–9369.

Duvernell DD, Eanes WF. 2000. Contrasting molecular population genetics of four hexokinases in *Drosophila melanogaster*, *D. simulans* and *D. yakuba*. Genetics. 156:1191–1201.

Eanes WF. 1999. Analysis of selection on enzyme polymorphisms. Annu Rev Ecol Syst. 30:301–326.

Eanes WF, Kirchner M, Yoon J. 1993. Evidence for adaptive evolution of the G6PD gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc Natl Acad Sci USA. 90:7475–7479.

Eanes WF, Kirchner M, Yoon J, Biermann CH, Wang I-N, McCartney MA, Verrelli BC. 1996. Historical selection, amino acid polymorphism and lineage-specific divergence at the G6pd locus in *Drosophila melanogaster* and *D. simulans*. Genetics. 144:1027–1041.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. Genetics. 162:2017–2024.

Gillespie J. 1991. The causes of molecular evolution. New York: Oxford University Press.

Grumbling G, Strelets V. The FlyBase Consortium. 2006. FlyBase: anatomical data, images and queries. Nucleic Acids Res. 34:D484–D488.

Hartl DL, Dykhuizen DE, Dean AM. 1985. Limits of adaptation: the evolution of selective neutrality. Genetics. 111:655–674.

Hartl DL, Taubes CH. 1996. Compensatory nearly neutral mutations: selection without adaptation. J Theor Biol. 182:303–309.

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

Hey J, Kliman RM. 2002. Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics. 160:595–608.

Hubbard TJP, Aken BL, Beal K, et al. (58 co-authors). 2006. Ensembl 2007. Nucleic Acids Res. 00:D1–D8.

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

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.

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

Labate J, Eanes WF. 1992. Direct measurement of *in vivo* flux differences between electrophoretic variants of *G6pd* from *Drosophila melanogaster*. Genetics. 132:783–787.

LaPorte DC, Walsh K, Koshland DE. 1984. The branch point effect: ultrasensitivity and subsensitivity to metabolic control. J Biol Chem. 259:4068–4075.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York: Columbia University Press.

Lewontin RC, Hubby JL. 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. 54:595–609.

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

Mitton J. 1997. Selection in natural populations. New York: Oxford University Press.

Moriyama EN, Powell JR. 1996. Intraspecific nuclear DNA variation in *Drosophila*. Mol Biol Evol. 13:261–277.

Oakeshott JG, Chambers GK, Gibson JB, Eanes WF, Willcocks DA. 1983. Geographic variation at *G6pd* and *Pgd* allele frequencies in *Drosophila melanogaster*. Heredity. 50:67–72.

Pamilo P, Bianchi NO. 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence of the genes. Mol Biol Evol. 10:271–281.

Pierce VA, Crawford DL. 1997. Phylogenetic analysis of glycolytic enzyme expression. Science. 276:256–259.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 16:266–274.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. Genetics. 132:1161–1176.

Sezgin E, Duvernell DD, Matzkin LM, Duan YH, Zhu CT, Verrelli BC, Eanes WF. 2004. Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. Genetics. 168:923–931.

Shapiro JA, Huang W, Zhang C, et al (12 co-authors) 2007. Adaptive genic evolution in the Drosophila genomes. Proc Natl Acad Sci USA. 104:2271–2276.

Sharp PM, Lloyd AT. 1993. Codon usage in *Drosophila* genes. In: Maroni G, editor. An atlas of Drosophila genes. New York: Oxford University Press. p. 378–397.

Stephanopoulos G, Vallino JJ. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science. 252:1675–1681.

Verrelli BC, Eanes WF. 2000. Extensive amino acid polymorphism at the *Pgm* locus is consistent with adaptive protein evolution in *Drosophila melanogaster*. Genetics. 156:1737–1752.

Verrelli BC, Eanes WF. 2001a. Clinal variation for amino acid polymorphisms at the *Pgm* locus in *Drosophila melanogaster*. Genetics. 157:1649–1663.

Verrelli BC, Eanes WF. 2001b. The functional impact of *Pgm* amino acid polymorphism on glycogen content in *Drosophila melanogaster*. Genetics. 159:201–210.

Vitkup D, Kharchenko P, Wagner A. 2006. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 7:R39.

Watt WB, Dean AM. 2000. Molecular functional studies of adaptive genetic variation in prokaryotes and eukaryotes. Annu Rev Genet. 34:593–622.

Wright F. 1990. The effective number of codons used in a gene. Gene. 87:23–29.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–505.