Amino Acid Polymorphism and Rare Electrophoretic Variants of G6PD From Natural Populations of *Drosophila melanogaster*

Walter F. Eanes, Michele Kirchner, Daniel R. Taub, Jeanne Yoon and Jiang-tian Chen[1]

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

Manuscript received February 15, 1993
Accepted for publication February 14, 1996

ABSTRACT

Identifying the amino acid changes responsible for electrophoretic variants is essential to understanding the significance of allozyme polymorphism in adaptation. The amino acid mutations responsible for the common G6PD allozyme polymorphisms in *Drosophila melanogaster* have been recently described. This study characterizes the amino acid changes associated with 11 rare electrophoretic G6PD variants. The 11 rare electrophoretic variants result from six independent amino acid mutations. The *in vivo* function of the rare variants was determined in an earlier study and most variants fell into one of two function classes. It is shown here that the function of the rare variants reflects the state of the Pro/Leu mutation responsible for the *A/B* allozyme polymorphism in each variant. Two mutations destabilize quaternary structure resulting in shifts from tetrameric to dimeric alleles, and one of these also results in a variant with *in vivo* function intermediate to *A* and *B*. That mutation is an aspartic-acid-to-asparagine change that is two residues away from the Pro/Leu polymorphism responsible for the *A/B* dimer-tetramer quaternary shift. Structure-function relationships based on studies of human G6PD deficency-associated mutations predict that these last two amino acid changes fall within the protein domain responsible for NADP binding.

THE description, more than a quarter of a century ago, of electrophoretic variation in Drosophila and humans (HARRIS 1966; HUBBY and LEWONTIN 1966) not only established the fundamental observation of widespread molecular polymorphism in natural populations, but also opened the field of population genetics to a wider audience. Population biologists investigating species untractable to genetic analysis, and lacking conspicuous visible polymorphisms, began to focus on population genetic questions. Central to the use of electrophoretic variation in studies of population structure and questions of its significance as adaptive variation has been an interest in the molecular nature of electrophoretic polymorphism. It has been accepted that electrophoretic variants are caused by underlying amino acid differences, but it is not apparent if electrophoretic "alleles" generally result from single or multiple substitutions, to what extent electrophoretic classes have multiple origins, sharing only net charge (see RAMSHAW *et al.* 1979), and which amino acid polymorphisms affect enzyme function.

The *G6pd* locus in *Drosophila melanogaster* provides an excellent model for the study of function and polymorphism in natural populations. G6PD has an explicit and simple function at the head of the pentose shunt and shows unambiguous conservation of structure across taxa (JEFFREY *et al.* 1993). In Drosophila the function of G6PD mutations can be assessed by either direct flux studies using radio-labeled glucose (CAVENER and CLEGG 1981; LABATE and EANES 1992), indirect assay using the strong viability interaction between activities between G6PD and 6-phosphogluconate dehydrogenase, or 6PGD (EANES 1984; EANES and HEY 1986; EANES *et al.* 1990).

The electrophoretic polymorphism have been the focus of considerable study in *D. melanogaster* (see for example, YOUNG *et al.* 1964; CAVENER and CLEGG 1981; GANGULY *et al.* 1985; EANES *et al.* 1989; MIYASHITA 1990). The cosmopolitan diallele polymorphism is widespread, showing reciprocating latitudinal clines in northern and southern hemispheres (OAKESHOTT *et al.* 1983). The *A* allele, which results from a Pro to Leu mutation at residue 382 in all copies sampled so far (EANES *et al.* 1993), migrates as a dimer under electrophoresis, and predominates in Europe and Japan, while the tetrameric *B* variant predominates in sub-Saharan Africa. An earlier search for cryptic electrophoretic variation (EANES 1983) focused on this question and turned up no evidence for hidden molecular variation, but without direct sequencing, the issue remained open until recently.

Our studies of *in vitro* and *in vivo* function clearly show that the common electromorphs are associated with different pentose shunt function, with the *A* variant possessing 20–40% lower activity (EANES *et al.* 1990; LABATE and EANES 1992). Furthermore, an earlier study of the *in vivo* function associated with a set of 11 rare

Corresponding author: Walter F. Eanes, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794. E-mail: walter@life.bio.sunysb.edu

[1]*Present address:* Group in Biostatistics, 140 Warren Hall, University of California, Berkley, CA 94720.

G6PD electrophoretic variants recovered from a screen of over 15,000 chromosomes indicated two clusters of G6PD function, and we proposed without any knowledge of the underlying amino acid changes, that this reflected the common allele from which each rare variant was derived (EANES and HEY 1986). This report addresses several questions. What are the types of amino acid changes associated with the 11 rare G6PD variants seen in that study? Are the rare variants showing shared electrophoretic mobility also identical by descent? Are most rare variants newly derived? What is the relationship between the amino acid changes and their relationship to *in vivo* function and structure-function features as ascertained from studies of mutation in the human G6PD. Are the amino acid mutations associated with polymorphisms fundamentally different than those associated with interspecific divergence?

MATERIALS AND METHODS

Wild and mutant lines: The origin and genetic extraction of *X* chromosomes for the 13 rare *G6pd* variants from *D. melanogaster* are described in EANES and HEY (1986).

PCR amplification and DNA sequencing: DNA sequence representing the 1673 nucleotide segment of the *G6pd* locus (545–2216 in FOUTS *et al.* 1988) in *D. melanogaster* was amplified via PCR from genomic DNA prepared as in MCGINNIS *et al.* (1983). Approximately 10 ng of genomic DNA was amplified in 10 mM Tris (pH 8.3), 50 mM KCl, 0.01% gelatin, 1 mM $MgCl_2$, 2 units of *AmpliTaq* polymerase (Perkin-Elmer), and 100 nM of each primer. The resulting amplified 1.67-kb fragment was then excised from 3% NuSieve agarose and used as template to amplify two smaller segments (545–1436 and 1408–2216). Single strand template for sequencing was generated by the kinased primer/λ exonuclease digestion described by HIGUCHI and OCHMAN (1989). DNA template was separated from PCR primers using Millipore filters. Primers for Sanger dideoxy sequencing using Sequenase (U.S. Biochemical) were spaced about every 300 bp. Each sequencing reaction was run on both standard acrylamide gels with an electrolyte gradient (SHEEN and SEED 1988), and Long Ranger acrylamide gels (AT Biochem). Both strands were completely sequenced for each allele, with rare gaps of 5–10 bases (<1–2% of the total sequence), where only one strand produced readable sequence. All polymorphisms were confirmed on both strands, and no errors were observed. The 1706-base sequence for the 11 gene copies described here are stored under EMBL Accession numbers Z19019-Z19021, Z190030-Z19034, Z19042, Z19043 and Z19048.

RESULTS AND DISCUSSION

Table 1 lists the 11 nucleotide changes that result in amino acid replacements for the rare *G6pd* alleles. Among the 11 variants there are six replacement events. It would appear that alleles *AF2, AF3* and *AF4* are all identical-by-descent, as are alleles *AS1, AS3, AS4* and *AS5*. Allele *AS2* is an electrophoretic variant that migrates as a dimer, yet is derived from the *B* allele, and results from a substitution of asparagine for aspartic acid at residue 384, which is only two residues from the *A/B* proline-leucine change. In addition to this sequence of exons 2–4, we have sequenced another 900 bases around exon 1 (coding for six amino acids) in *Nash*$^-$, but failed to identify any amino acid or other change responsible for activity loss of that allele. Earlier studies showed that the *Blo1* low activity allele is the result of a *P*-element insertion in the first intron and this allele was not sequenced (EANES *et al.* 1990).

In our earlier work on *in vivo* and *in vitro* function associated with the G6PD polymorphism, we assumed that any catalytic differences responsible for the activity differences were the consequence of a single amino acid difference. This was because a screen of 126 lines from France and North America, using two different pHs and varying acrylamide concentrations, failed to find any hidden molecular variation within either electrophoretic class (EANES 1983). Consistent with that observation, we have reported that the common electrophoretic polymorphism for dimeric *A* and tetrameric *B* alleles is the result of a single change, a proline-to-leucine substitution at amino acid 382 (EANES *et al.* 1993). We proposed that this single amino acid change accounts for the *A/B* allozyme polymorphism (the quaternary structure difference) and the catalytic differences reported in EANES *et al.* (1990).

The G6PD rare variants described in EANES and HEY (1986) presented a unique opportunity to further investigate the relationship between amino acid substitution and function at this gene, albeit they are not a random sample of mutation, but those mutations which have survived to be sampled in natural populations. In that study, we used the epistatic interaction between G6PD and 6PGD activity to assess *in vivo* function for the 11 rare variants sequenced here. The experimental design uses *X* chromosome constructions where each *G6pd* allele is placed in a low activity 6PGD background provided by the $6Pgd^{lo1}$ allele. Complete null alleles of 6PGD are lethals, but this leaky allele exhibits viability variation that is dependent on the activity of *G6pd* alleles preceding it in the pentose shunt; the lower the G6PD activity, the greater the suppression of lethality associated with $6Pgd^{lo1}$ (EANES 1984; EANES and HEY 1986; EANES *et al.* 1990). In this scheme *A* genotypes almost fully suppress $6Pgd^{lo1}$ lethality, while *B* genotypes do poorly. In our study of the *in vivo* function associated with a set of 11 rare G6PD electrophoretic variants recovered from natural populations, we observed two general clusters of G6PD function. Consequently, we proposed, without any knowledge of the underlying amino acid changes, that this reflected the common allele from which each rare variant was derived (EANES and HEY 1986).

In the present study, we observed that the 11 variants are the result of six independent mutation events, with two events appearing derived from the *A* allele sequences (*AF1* and *AF2-4*), and four from the *B* allele (some alleles are identical-by-descent). As predicted in EANES and HEY (1986), the dichotomous *in vivo* func-

TABLE 1

G6PD variants and associated mutations

Allele	Nucleotide[a]	Replacement	*A* or *B*[b]	Locality
A	1817 C → T	382 Pro → Leu		See text
AF1	619 G → T	32 Gly → Cys	*A*	Europe
AF2	942 C → T	119 Arg → Cys	*A*	Mt. Sinai, NY
AF3	942 C → T	119 Arg → Cys	*A*	St. James, NY
AF4	942 C → T	119 Arg → Cys	*A*	Lincoln, MA
AS1	1349 G → A	226 Ser → Asn	*B*	Ft. Pierce, FL
AS2	1822 G → C	384 Asp → Asn	*B*	Ft. Pierce, FL
AS3	1349 G → A	226 Ser → Asn	*B*	Orchid, FL
AS4	1349 G → A	226 Ser → Asn	*B*	Orchid, FL
AS5	1349 G → A	226 Ser → Asn	*B*	Alexander Springs, FL
BS1	1517 G → A	282 Ser → Asn	*B*	Somerville, MA
BF1	928 G → A	114 Gly → Asp	*B*	Watsonville, CA
Nash$^-$	None	None	*B*	Nashville, TN

List of the 12 rare G6PD variants sequenced, the mutations responsible for the amino acid changes presumed associated with each mobility variant, with the exception of *Nash*$^-$, which was recovered as a null allele. The state of each variant at residue 382 (*A* or *B*) is also given.

[a] Polarity based on *Drosophila simulans* sequence (EANES *et al.* 1993).

[b] Leucine/proline polymorphism, *A*, Leu; *B*, Pro.

tion (inversely related to "viability depression"; Figure 1) seen for the collection of variants largely reflects the state of the leucine-proline polymorphism at residue 382 (Table 1). The exceptions are *Nash*$^-$, whose low activity is not due to any amino acid change, and B^{lo1}, whose low activity results from a *P*-element insertion in intron 1 (EANES *et al.* 1990). The *AS2* allele retains the proline at position 382, yet migrates as a dimer and possesses a radical substitution of asparagine for aspartic acid at residue 384. Thus, this mutation is only two amino acid residues removed from the proline-leucine polymorphism responsible for the dimer-tetramer transition seen in *A*/*B*. It also shows *in vivo* function intermediate to *A* and *B*. This implicates this part of the amino acid sequence with dimer-dimer associations, and the catalytic differences seen between *A* and *B* variants as well.

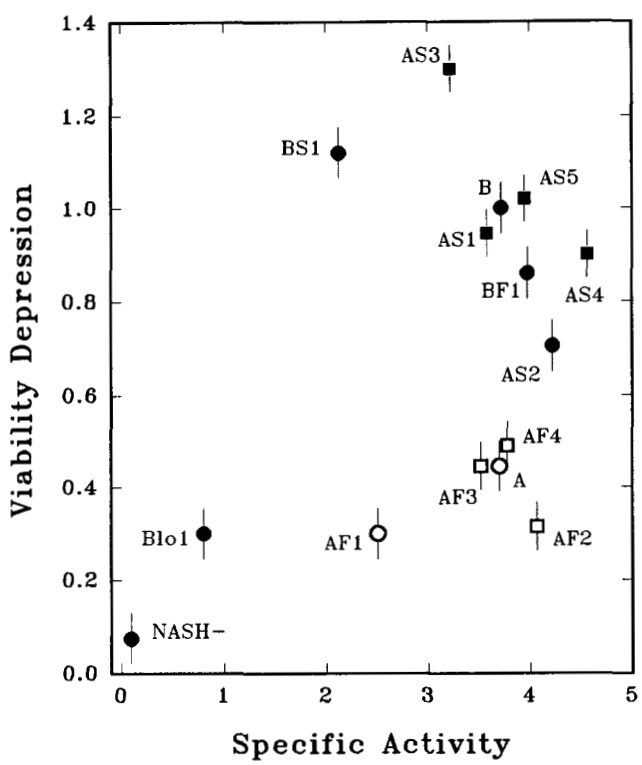

FIGURE 1.—The relationship between standardized viability depression, as described in EANES and HEY (1986), for males carrying the $6Pgd^{lo1}$ allele and each *G6pd* variant and measured larval-specific crude activities. The solid and filled symbols represent alleles carrying a proline (*B* allele) or leucine (*A* allele) at residue 382, respectively. The squares represent two sets of alleles with identical mutations.

We might expect most allozyme polymorphisms, if effectively neutral, to involve amino acid changes that are benign in the sense that they substitute residues of similar physiochemical nature and involve protein domains of little importance. A proline-to-leucine change, as seen for the *A*/*B* polymorphism is not functionally conservative. In studies of conformational stability, proline is singled out for its key role in chain conformation, protein folding, and stability, and there are numerous examples of Pro ↔ Leu mutations that yield unstable proteins (PAKULA and SAUER 1989; MACARTHUR and THORNTON 1991). Consistent with this the *A* enzyme is significantly more thermolabile (BIJLSMA and VAN DER MEULEN-BRUIJNS 1979).

Another electrophoretic variant, designated *AF1*, is polymorphic in European populations and is the result of a Gly-to-Cys substitution in exon 2 (EANES *et al.* 1993, 1995). This allele is derived from the *A* allele sequence, possesses slightly lower *in vivo* function, and is even more thermolabile than the *A* enzyme (EANES and HEY 1986). This is another example of a radical amino acid change that contributes to a high-frequency polymor-

phism. Glycine may assume many backbone angles energetically unfavorable to other amino acid residues, and substitutions are often destabilizing (PAKULA and SAUER 1989).

Although the alleles *AF2, AF3* and *AF4* are identical in nucleotide sequence, they were independently recovered from collections from St. James and Mt. Sinai, New York (Long Island), and Lincoln, Massachusetts, in 1981 and 1982. The associated mutation of Arg to Cys at residue 119 should change the net charge to one more negative and result in the observed faster electrophoretic mobility.

In principle, the dimer-tetramer transition that characterizes the *A/B* polymorphism could be generated by any number of amino acid mutations, and an early, albeit unrealized, concern of population level studies was that the *A* electromorph would be a class of heterogeneous mutations. In this study, two other mutations were identified that also cause disruption of quaternary structure leading to dimeric alleles. The *AS2* allele resulting from a change of Asp to Asn at 384 and the Ser-to-Asn change at residue 226 that is collectively responsible for alleles *AS1, AS3, AS4* and *AS5* are both associated with dimeric variants. All were collected from widely separated sites in Florida in 1983. The latter set has a quaternary structure that, upon electrophoresis, is pH dependent; the variants carrying this change migrate as a slow dimer at pH 8.9 and migrate as both a slightly fast tetramer and slow dimer at pH 7.0.

None of the rare sequences possess unique synonymous polymorphisms which might be indicative that they are were old. Our knowledge of the of the *D. simulans* and *D. yakuba* sequences shows that all the rare variants, as well as the common *A* and *AF1* alleles, are derived; none represent an ancestral allele that is now near extinction.

Additional insight into the structure-function relationships of G6PD variants in Drosophila can be obtained by comparison with human G6PD and its many deficiency mutations. More than 40 variants of human G6PD have been sequenced and characterized (BEUTLER 1991; VULLIAMY *et al.* 1992). In general, these variants are not randomly sampled, but represent alleles deficient for G6PD activity, a condition that ~400 million of the world human population carry in response to malarial selection (see review by VULLIAMY *et al.* 1992). Figure 2 shows the distribution of 33 amino acid mutations that have been identified as deficient for human red cell G6PD. They are also characterized with respect to extent of the G6PD deficiency. The majority have been identified because they cause loss of red cell G6PD activity, but do not result in sufficient loss of activity in other tissues, which is assumed to be a lethal phenotype. Thus, this the distribution plot of a particular type of structure-function disruption. The figure also shows those residues that have remained unchanged across the yeast-human-rat-Drosophila lineages (PERSSON *et al.* 1991). Two blocks of amino acids, from residues 188 to 215 (which includes the proposed reactive Lys 204) and 372 to 413, have been proposed as involved in glucose-6-phosphate and NADP binding, respectively (HIRONO *et al.* 1989; BEUTLER 1991), and it is apparent that these regions are where the majority of deficency-associated mutations reside. There is also a highly conserved central core from residues 224 to 292, and mutations to deficiencies apparently do not survive if associated with this region.

It is very interesting that both the Pro/Leu substitution at 382 and the rare variant *AS2* 384 fall within the block of amino acid residues (372–413 in human sequence) characteristically associated with G6PD-derived severe nonspherocytic anemia in humans. This domain is associated with NADP binding (HIRONO *et al.* 1989; BEUTLER 1991), although it has been questioned if this is the binding site *per se* (JEFFREY *et al.* 1993). Dimer-dimer association and the stability of the *A* variant are known to depend on NADP/NADPH environment (WILLIAMSON and BENTLEY 1983), and the *A* and *B* variants differ significantly in their K_Ms for NADP as well as glucose-6-phosphate (human deficiency mutations typically show modifications of both K_Ms). Overall, this analysis or rare variants underscores the functional impact of the proline/leucine mutation that is responsible for the *A/B* electrophoretic polymorphism. A couple of other mutations have changes worth noting. Both *BF1* and *BS1* involve amino acids that have remained unchanged across distant taxa. *BF1* shows no loss of function as measured crudely *in vitro* or *in vivo,* but lies in a rather benign region of the enzyme (EANES and HEY 1986). In contrast, *BS1* shows a conservative change, but in an invariant residue. It is worth noting that this variant has normal *in vivo* function, but is nevertheless very unstable.

For contrast with the observed common polymorphisms and rare variants, we have also plotted the 24 amino acid mutations that have become fixed in the *D. melanogaster-simulans-yakuba* lineages (EANES *et al.* 1993, 1995). Consistent with the general principles of functional constraint and molecular evolution (KIMURA 1983), the observed amino acid differences between species are found almost exclusively in regions of high amino acid divergence and outside the regions with high densities of deficiency mutations (Figure 2). Two rare mutants recovered in our survey (*BF1* and the set *AF2-AF4*), both showing no gross changes in *in vivo* function, are also found in the span of residues from 76 to 126 (in human sequence), which constitutes only 9.5% of the amino acid sequence, yet it is where 14 of 24 (or 58%) of the fixed differences appear.

We might also expect fixed interspecific differences to be of a functionally more conservative type of transition that polymorphic mutations. This is difficult to examine in an objective fashion, but measures of physiochemical difference between amino acids have been

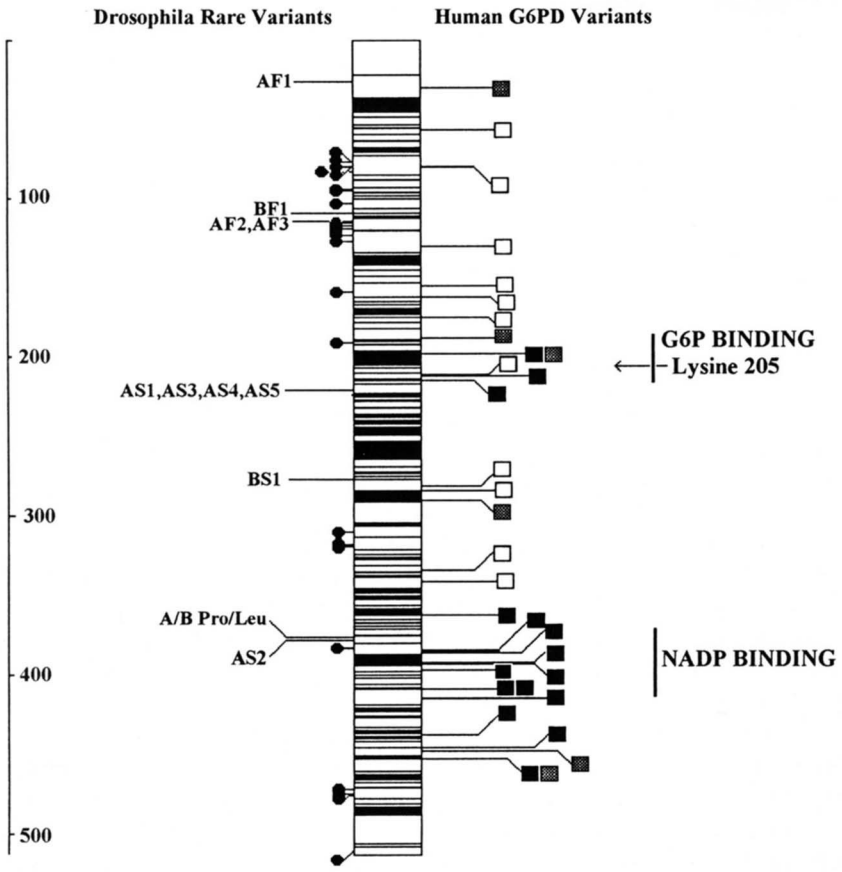

FIGURE 2.—Position of the amino acid mutations found for the 11 rare and two common *G6pd* variants of *D. melanogaster* relative to the position of 33 human G6PD deficiency mutations (right). The cross banding designates residues that are unchanged in the yeast, rat, human, and *D. melanogaster* sequences. The extent of activity deficiency of the human mutations are given by the symbols; ■, nonspherocytic hemolytic anemia; ▨, severe enzyme deficiency; and □, moderate deficiency; see BEUTLER (1991).

proposed by several investigators (SNEATH 1966; GRANTHUM 1974; MIYATA *et al.* 1979) and all three measures are correlated with substitution rate. The MIYATA *et al.* (1979) physiochemical distance varies from a low of $X = 0.06$ for the Pro-Ala pair to $X = 5.13$ for the Trp-Gly pair, with a mean of 1.93 based on KIMURA's (1983) analysis. We have computed the average MIYATA's *et al.* distance for the 19 fixed mel-sim amino acid differences and seven amino acid polymorphisms, and they are 0.98 and 1.95, respectively. This is a significant difference when tested by the nonparametric Wilcoxon two-sample test ($U_S = 109$; $0.02 > P > 0.01$; SOKAL and ROHLF 1969). Therefore, we also conclude that amino acid polymorphisms involve more radical types of physiochemical change.

If natural selection favors differential partitioning of pathway fluxes in different environments (*i.e.*, lower activity has higher fitness in some niches) then sufficient changes in enzyme activities may only be brought about by less conservative amino acid changes. However, the observation that polymorphisms are more physiochemically radical than fixed differences between species does not indicate that these are adaptive polymorphisms. OHTA's (1973) theory of nearly neutral mutation predicts that slightly deleterious alleles will contribute significantly more to intraspecific polymorphism than to fixed substitutions between species. It is expected that amino acid mutations arising as fixed differences between species will involve more conservative changes in more benign regions of the enzyme than polymorphisms. A surprising number of polymorphisms that have been characterized in detail show substantial differences for *in vitro* and *in vivo* function (WATT 1994), and radical amino acid changes in important functional domains may be a characteristic of such polymorphisms. Of course, caution must be exercised in anecdotal mutation-by-mutation inference. General statements about the types of mutations involved in polymorphisms can only emerge from a large body of overall pattern, which will emerge once more and more polymorphisms are examined with the advent of large scale sequencing.

We thank MARTY KREITMAN for introducing us to PCR-faciltated DNA sequencing. Thanks are extended to CEDRIC WESLEY in the early phases of the project and to JOANNE LABATE and DAN DYKHUIZEN for reading early versions of the manuscript. This study was supported, in part, by grants BSR-907391 from the National Science Foundation and GM-45247 from the National Institutes of Health to W.F.E. This is contribution number 856 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

LITERATURE CITED

BEUTLER, E., 1991 Glucose-6-phosphate dehydrogenase deficiency. N. Engl. J. Med. **324:** 169–174.

BIJLSMA, R., and C. VAN DER MEULEN-BRUIJNS, 1979 Polymorphism at the G6PD and 6PGD loci in *Drosophila melanogaster*. III. Developmental and biochemical aspects. Biochem. Genet. **17:** 1131–1144.

CAVENER, D. R., and M. T. CLEGG, 1981 Evidence for biochemical and physiological differences between genotypes in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **78:** 4444–4447.

EANES, W. F., 1983 Genetic localization and sequential electrophoresis of G6pd in *Drosophila melanogaster*. Biochem. Genet. **21:** 703–711.

EANES, W. F., 1984 Viability interactions, *in vivo* activity and the G6PD polymorphism in *Drosophila melanogaster*. Genetics **106:** 95–107.

EANES, W. F., and J. HEY, 1986 *In vivo* function of rare G6pd variants from natural populations of *Drosophila melanogaster*. Genetics **113:** 679–693.

EANES, W. F., J. W. AJIOKA, J. HEY and C. WESLEY, 1989 Restriction map variation associated with the G6PD polymorphism in *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 384–397.

EANES, W. F., L. KATONA and M. LONGTINE, 1990 Comparison of *in vitro* and *in vivo* activities associated with the G6PD allozyme polymorphism in *Drosophila melanogaster*. Genetics **125:** 845–853.

EANES, W. F., M. KIRCHNER and J. YOON, 1993 Evidence for adapative evolution of the G6PD gene in the *Drosophila melanogaster* and *D. simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

FOUTS, D., R. GANGULY, A. G. GUTIERREZ, J. C. LUCCHESI and J. E. MANNING, 1988 Nucleotide sequence of the Drosophila glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene. Gene **63:** 261–275.

GANGULY, R., N. GANGULY and J. E. MANNING, 1985 Isolation and characterization of the glucose-6-phosphate dehydrogenase gene in *Drosophila melanogaster*. Gene **35:** 91–101.

GRANTHAM, R., 1974 Amino acid distance formula to help explain protein evolution. Science **185:** 862–864.

HARRIS, H., 1966 Enzyme polymorphism in man. Proc. R. Soc. Lond. B Biol. Sci. **164:** 298–310.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HIRONO, A., W. KUHL, T. GELBART, L. FORMAN, V. F. FAIRBANKS *et al.*, 1989 Identification of the binding domain for NADP of human glucose-6-phosphate dehdrogenase by sequence analysis of mutants. Proc. Natl. Acad. Sci. USA **86:** 10015–10017.

HUBBY, J. L., and R. C. LEWONTIN, 1966 A molecular aproach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics **54:** 577–594.

JEFFERY, J., B. PERSSON, I. WOOD, T. BERGMAN, R. JEFFERY *et al.*, 1993 Glucose-6-phosphate dehydrogenase: structure-function relationships and the *Pichia jadinii* enzyme structure. Eur. J. Biochem. **212:** 31–49.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

LABATE, J., and W. F. EANES, 1992 Direct measurement of *in vivo* flux differences between electrophoretic variants of G6PD from *Drosophila melanogaster*. Genetics **132:** 783–787.

MACARTHUR, M. W., and J. M. THORTON, 1991 Influence of proline residues on protein conformation. J. Mol. Biol. **218:** 397–412.

MCGINNIS, W., A. W. SHERMOEN and S. K. BECKENDORF, 1983 A transposable element inserted just 5′ to a Drosophila glue gene alters gene expression and chromatin structure. Cell **34:** 75–84.

MIYASHITA, N. T., 1990 Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. Genetics **125:** 407–419.

MIYATA, T., S. MIYAZAWA and T. YASUNAGA, 1979 Two types of amino acid substitution in protein evolution. J. Mol. Evol. **12:** 219–236.

OAKESHOTT, J. G., G. K. CHAMBERS, J. B. GIBSON, W. F. EANES and D. A. WILLCOCKS, 1983 Geographic variation in G6pd and Pgd allele frequencies in *Drosophila melanogaster*. Heredity **50:** 67–72.

OHTA, T., 1973 Slightly deleterious mutatnt substitutions in evolution. Nature **246:** 96–98.

PAKULA, A. A., and R. T. SAUER, 1989 Genetic analysis of protein stability and function. Annu. Rev. Genet. **23:** 289–310.

PERSSON, B., H. JÖRVALL, I. WOOD and J. JEFFREY, 1991 Functionally important regions of glucose-6-phospahte dehydrogenase defined by the *Saccharomyces cerevisiae* enzyme and its differences from the mammalian and insect forms. Eur. J. Biochem. **198:** 485–491.

RAMSHAW, J. A. M., J. A. COYNE and R. C. LEWONTIN, 1979 The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics **93:** 1019–1037.

SHEEN, J., and B. SEED, 1988 Electrolyte gradient gels for DNA sequencing. Biotechniques **6:** 942–944.

SNEATH, P. H. A., 1966 Relationships between chemical structure and biological activity in peptides. J. Theor. Biol. **12:** 157–195.

SOKAL, R. R., and F. J. ROHLF, 1969 *Biometry*. W. H. Freeman and Co., San Francisco.

VULLIAMY, T., P. MASON and L. LUZZATTO, 1992 The molecular basis of glucose-6-phosphate dehydrogenase deficiency. Trends Genet. **8:** 138–143.

WATT, W. B., 1994 Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool? Genetics **136:** 11–16.

WILLIAMSON, J. H., and M. M. BENTLEY, 1983 Comparative properties of three forms of glucose-6-phosphate dehydrogenase in *Drosophila melanogaster*. Biochem. Genet. **21:** 1153–1166.

YOUNG, W. J., J. E. PORTER and B. CHILDS, 1964 Glucose-6-phosphate dehydrogenase in Drosophila: X-linked electrophoretic variants. Science **143:** 140–141.

Communicating editor: A. G. CLARK