Copyright © 1994 by the Genetics Society of America

Evolutionary Inferences From DNA Variation at the 6-Phosphogluconate Dehydrogenase Locus in Natural Populations of Drosophila: Selection and Geographic Differentiation

David J. Begun and Charles F. Aquadro

Section of Genetics and Development, Biotechnology Building, Cornell University, Ithaca, New York 14853

Manuscript received October 25, 1991
Accepted for publication September 7, 1993

ABSTRACT

Several allozyme-coding genes in *Drosophila melanogaster* show patterns suggesting that polymorphisms at these loci are targets of balancing selection. An important question is whether these genes have similar distributions of underlying DNA sequence variation which would indicate similar evolutionary processes occurring in this class of loci. One such locus, 6-phosphogluconate dehydrogenase (*Pgd*), has previously been shown to exhibit clinal variation for Fast/Slow electromorph variation in the United States and Australia, unusually large electromorph frequency differences between the United States and Africa, and other patterns indicative of selection. We measured four-cutter DNA restriction site and allozyme variation at *Pgd* among 142 *D. melanogaster X* chromosomes collected from several geographic regions including North Carolina, California, and Zimbabwe (Africa). We also sequenced a representative sample of 13 *D. melanogaster Pgd* genes collected in North Carolina and a single copy of *Pgd* from the sibling species, *Drosophila simulans*. While some population genetic models predict excess DNA polymorphism in genes which are targets of balancing selection, the *D. melanogaster* samples from the United States had significantly reduced levels of DNA polymorphism and extraordinarily high levels of linkage disequilibrium, providing evidence of hitchhiking effects of advantageous mutants at *Pgd* or at linked sites. Therefore, while selection has probably influenced the distribution of DNA variation at *Pgd*, the precise nature of these selective events remains obscure. Since the *Pgd* region appears to have low rates of crossing over, the reduced level of variation at this locus supports the idea that recombination rates are important determinants of levels of DNA polymorphism in natural populations. Furthermore, while patterns of allozyme variation are very similar at *Pgd* and *Adh*, the DNA data show that the evolutionary histories of these genes are dramatically different. We observed extensive differences in the amount and distribution of variation in *D. melanogaster Pgd* samples from the United States and Zimbabwe which cannot be explained by differential selection on the Fast/Slow polymorphism in these two geographic regions. Thus, genetic drift among partially isolated populations has also been an important factor in determining the distribution of variation at *Pgd* in *D. melanogaster*. Finally, we assayed four-cutter variation at *Pgd* in a sample of 19 *D. simulans X* chromosomes and observed reduced levels of DNA variability and high levels of linkage disequilibrium. These patterns are consistent with predictions of some hitchhiking models.

THE application of molecular methods to population genetics has resulted in increased ability to test evolutionary hypotheses regarding the significance of naturally occurring genetic variation (reviewed by HUDSON 1990). In particular, we are now in a position to attempt to determine the relative importance of mutation, recombination, genetic drift and natural selection to the distribution of DNA variation in natural populations. The sampling strategy for selecting loci for study will depend on the question of interest. For example, several polymorphic allozyme loci in *Drosophila melanogaster* exhibit patterns of clinal variation which suggest that the variability is influenced by natural selection (SINGH and RHOMBERG 1987b). Molecular data may allow us to determine if these genes show similar patterns because they have similar selective histories (*e.g.*, HUDSON and KAPLAN 1988).

The 6-phosphogluconate dehydrogenase (*Pgd*) locus in *D. melanogaster* codes for an enzyme of the pentose-phosphate shunt and has been the subject of extensive biochemical and population genetic research documenting geographic and temporal variation in allozyme frequencies in natural populations (CAVENER and CLEGG 1981a; SINGH, HICKEY and DAVID 1982; OAKESHOTT *et al.* 1983; SINGH and RHOMBERG 1987a), changes in allozyme frequencies in response to selection in artificial populations (BIJLSMA and VAN DELDEN 1977; BIJLSMA and KERVER 1980; CAVENER 1983), and *in vitro* and *in vivo* bio-

[1] The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession number U02288.

chemical differences between allozymes (BIJLSMA and VAN DER MEULEN-BRUIJNS 1979; CAVENER and CLEGG 1981b). These patterns are often taken as evidence that amino acid variation at *Pgd* is visible to selection. We have undertaken a molecular analysis of *Pgd* with the hope of inferring the history of this locus in *D. melanogaster* and its sibling species, *D. simulans* (BEGUN and AQUADRO 1991).

Here we report the results of four-cutter restriction map surveys from the *Pgd* region in population samples of *D. melanogaster* from North Carolina, from two California populations, and from Zimbabwe, Africa, as well as a small sample from France and Southeast Asia. We also present direct DNA sequence data from a sample of *Pgd* genes from North Carolina. Finally, we present the sequence of the *Pgd* gene from *D. simulans* and the results of a four-cutter survey of *Pgd* in this species.

MATERIALS AND METHODS

***D. melanogaster* samples:** *North Carolina:* Isofemale lines were established from flies collected by C.F.A. at the Farmer's Market in Raleigh, North Carolina, in November 1984 and were subsequently maintained as vial cultures. In 1987 a single, randomly selected male from each isofemale line was used to make a homozygous *X* chromosome stock by means of the *FM7a* balancer.

California: The *FM7a* balancer was used to make homozygous *X* chromosome lines from the attached-*X* lines of the two California populations described in BEGUN and AQUADRO (1991).

Zimbabwe: Isofemale lines were established from flies collected by R. R. RAMEY and L. BROWN on September 26, 1990, at the Sengwa Wildlife Research Institute in Zimbabwe, Africa. Several male flies were also included in this collection. In October 1990 each wild-caught male, and a single male from each isofemale line, were used to make homozygous *X* lines with the *FM7a* balancer.

France and Southeast Asia: These isofemale lines are described in SINGH, HICKEY and DAVID (1982).

***D. simulans* sample:** A single male was selected from each of 19 of the inbred lines described in AQUADRO, LADO and NOON (1988) and crossed to an attached-*X* stock. DNA extracted from males from these attached-*X* lines was used for subsequent restriction map analysis. The *D. simulans* *Pgd* allele which was cloned and sequenced (see below) was from one of the aforementioned inbred lines.

Allozyme analysis: Several males from each line were homogenized manually in microfuge tubes in 1 × Tris-borate-EDTA buffer plus Ficoll loading dye. The resulting homogenate was electrophoresed for 3–4 hr at 100 V through 0.8% agarose, 1 × Tris-borate-EDTA gels at 4°. The gels were stained for PGD according to standard procedures (HARRIS and HOPKINSON 1976). We verified that electrophoretic variants observed using starch gel electrophoresis were scoreable using the above method.

Restriction map analysis: Approximately 3–4 μg of cesium chloride gradient-centrifugation-prepared DNA (BINGHAM, LEVIS and RUBIN 1981) from each line was digested with each of 10 four-base recognizing restriction enzymes (*Alu*I, *Dde*I, *Hae*III, *Hha*I, *Hin*fI, *Msp*I, *Rsa*I, *Sau*3AI, *Scr*FI and *Taq*I). The DNA was then ethanol precipitated, dried, and resuspended in 3 μl of formamide loading dye. Phage φX174 DNA cut with *Hae*III or with *Hin*fI was used as a size standard. These samples were heated to 95° and subjected to electrophoresis through 5% polyacrylamide gels (0.4 mm thick, 50 mM TBE, 8.3 M urea) at 55 W for approximately 2.5 hr (until the slower dye front was about 13 cm from the bottom of the gel). The anode and cathode contained 50 mM TBE and 100 mM TBE/1 M sodium acetate, respectively. DNA was electroblotted to Gene Screen membranes (NEN Research Products) and hybridized as described (KREITMAN and AGUADÉ 1986a). Membranes were probed with subclones or gel purified restriction fragments from *Pgd*7.4 (a pBR322 clone containing the entire *Pgd* gene provided by M. SCOTT and J. LUCCHESI). The probed region spans approximately 5.3-kb starting with the *Xho*I site at nucleotide position 1 of SCOTT and LUCCHESI (1991) and extending through the *Sal*I site (see Figure 3 of SCOTT and LUCCHESI 1991) at approximately nucleotide position 5300. Probes were radiolabeled with [^{32}P]dCTP by random priming (FEINBERG and VOGELSTEIN 1983).

Cloning and sequencing of the *Pgd* locus: We cloned a conserved 9-kb *Bgl*II fragment containing the entire *Pgd* gene (BEGUN and AQUADRO 1991) into the λZapII vector (Stratagene, Inc.). λZapII DNA was treated with T4 DNA ligase and then digested with *Xho*I. The *Xho*I termini were partially filled in with dTTP and dCTP by the Klenow fragment of DNA polymerase so as to be compatible with partially filled-in *Bgl*II termini. Genomic DNA (10–15 μg) from 13 *D. melanogaster* lines and 1 *D. simulans* line were digested to completion with *Bgl*II and subjected to electrophoresis through 0.8% low melting point agarose in 0.8 mM Tris, 0.4 mM acetic acid, and 0.04 mM EDTA, pH 8.0 buffer. DNA fragments ranging in size from about 8.5 to 9.5 kb were isolated from the gel and purified by phenol extraction and ethanol precipitation. The termini of these fragments were partially filled in with dGTP and dATP, phenol extracted, precipitated with ethanol and resuspended in a small volume of water. The partially filled-in genomic DNA and λZapII DNA were ligated with T4 DNA ligase and packaged using Gigapack (Stratagene, Inc.) *in vitro* packaging extracts. Approximately 50,000 pfu from each library were screened according to standard procedures (SAMBROOK, MANIATIS and FRITSCH 1989) with a 2.3-kb *Bam*HI/*Sal*I fragment containing the large exon of *Pgd*. Single plaques obtained after secondary screens were used to produce Bluescript clones using the *in vitro* excision process (Stratagene, Inc.); both strands of the *Pgd* region in the resulting double-stranded plasmid were sequenced using primers designed from the published *Pgd* sequence or from our own sequence data.

RESULTS

In a total of 142 *D. melanogaster X* chromosomes sampled from several geographic regions we observed three electromorphs, Slow, Fast, and the rare Very Fast. We observed 30 polymorphic restriction sites out of a total of 200 sites scored (Table 1; Figure 1). Direct sequencing of 13 *D. melanogaster Pgd* genes collected in North Carolina revealed 17 substitutions out of 4203 bases surveyed (Table 2). The Fast/Slow polymorphism results from a C/A (glutamine/lysine) change at nucleotide position 3317 and is the only replacement polymorphism among the 13 genes. We scored 180 four-cutter restriction sites in *D. simulans*,

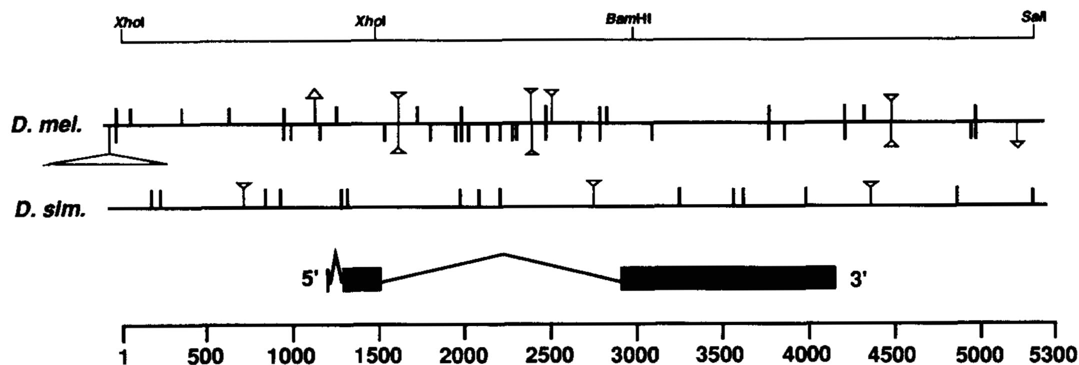

FIGURE 1.—Polymorphisms in the *Pgd* region of *D. melanogaster* and *D. simulans*. The top line represents the region spanned by the probe. For *D. melanogaster*, sites which were polymorphic only in Zimbabwe are below the line, sites which were polymorphic in other samples are indicated above the line, and sites polymorphic in both Zimbabwe and other samples are shown above and below the line. Size variants in both species are indicated by open triangles.

15 of which were polymorphic (Table 3; Figure 1). All size variation in both species was in non-coding DNA. Allozyme variation was not measured for *D. simulans*, as the lines were no longer available.

Comparison of the DNA sequence of *Pgd* in *D. melanogaster* and *D. simulans*: In Figure 2 we show the DNA sequence of the *D. simulans Pgd* region. The sequence starts at nucleotide position 466 of SCOTT and LUCCHESI (1991). In the process of collecting the data for this report we discovered that the SCOTT and LUCCHESI (1991) sequence was inconsistent with some of our results. MAX SCOTT (personal communication) informed us that the SCOTT and LUCCHESI (1991) sequence can be corrected by substituting nucleotides 1727–1960 and 2688–2800 by their reverse-complement. Locations of polymorphisms within *D. melanogaster* and *D. simulans* and comparison between species are with reference to this corrected sequence. There were two nucleotide differences between our Slow *Pgd* genes and that of SCOTT and LUCCHESI (1991). We found a G at nucleotide 2880 where they report a C; we found a G inserted between nucleotides 493 and 494. Both changes match the *D. simulans* sequence and occur in non-coding DNA. A summary of sequence differences between species is shown in Table 4.

Population differentiation in *D. melanogaster*: We used the methods of HUDSON, BOOS and KAPLAN (1992) to compare populations using four-cutter restriction site haplotypes (size variants not included; Table 5). We estimated F_{ST} using equation 3 of HUDSON, SLATKIN and MADDISON (1992) with estimates of within population variability weighted by sample size. The Arvin, California, and Soda Lake, California, samples were not significantly different and were pooled (subsequently referred to as the California population). The North Carolina and California populations are distinct when all chromosomes are included. When Slows are eliminated from the analysis there is still some evidence that the populations are different.

The Zimbabwe sample is very different from the U.S. samples. Only one Slow electromorph (haplotype 19) was observed among the 48 sampled chromosomes. This is consistent with previous allozyme surveys of African populations which indicated a low frequency of the Slow electromorph in Africa (SINGH, HICKEY and DAVID 1982; SINGH and RHOMBERG 1987a). A parsimony analysis shows that this haplotype does not belong in the same clade as the U.S. Slows (data not shown), though there were not enough phylogenetically informative sites to determine whether the allele represents an independent origin of the Slow electromorph or is instead, an old allele from the same lineage as the U.S. Slows. Zimbabwe and America share no four-cutter haplotypes (Table 1). When size variation is excluded, only two of the 20 Zimbabwe haplotypes are found in America. Comparisons of the Zimbabwe and U.S. samples using the permutation procedures of HUDSON, BOOS and KAPLAN (1992) were highly significant for the entire data set and when U.S. Slows were eliminated from the analysis (Table 5), suggesting that different frequencies of electromorphs can not account for the population heterogeneity.

Linkage disequilibrium: In *D. melanogaster*, there was a very high level of linkage disequilibrium among variants within the approximately 5.3-kb region surveyed with four-cutters. In North Carolina, 11 variants occurred with frequency greater than 0.1. All 55 pairwise comparisons among these sites had a value of D' (LEWONTIN 1964) significantly different from zero ($P < 0.05$, Fisher's exact test). Fifty-two of the 55 comparisons had a D' of +1 or −1, indicating the presence of only two or three gametic types. Three pairwise comparisons showing four gametic types (*Rsa*I 927 *vs.* *Hha*I 3784, Fast/Slow, or indelD) are attributable to haplotype 27, which is probably a recombinant between Fast and Slow haplotypes. Direct sequence analysis of 13 *Pgd* genes from North Carolina revealed similar levels of linkage disequilibrium. In California, 11 variants occurred with fre-

TABLE 1

Four-cutter restriction map variation at the *Pgd* locus of *D. melanogaster*

	Haplotype																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 *Alu*I −29[a]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−
2 *Hha*I 55[a]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
3 *Scr*FI 378 (T/C)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4 *Alu*I 612–615	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5 *Rsa*I 927 (C/T)	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
6 *Msp*I 980–983	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+
7 *Alu*I 1162–1165	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+
8 *Scr*FI 1241–1246	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
9 *Taq*I 1529 (C/T)	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
10 *Scr*FI 1721 (G/C)	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
11 *Alu*I 1807–1810	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	+	+	+
12 *Taq*I 1939–1942	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	+	+	+	+	+	+	−
13 *Taq*I 1954 (T/C)	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
14 *Msp*I 1972 (C/G)	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
15 *Sau*3A 1999–2003	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
16 *Scr*FI 2116–2120	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+
17 *Sau*3A 2175–2178	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
18 *Hae*III 2270 (G/C)[a]	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
19 *Hin*fI 2288 (T/C)	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
20 *Hin*fI 2474 (A/G)[a]	+	+	+	+	+	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+
21 *Hae*III 2639–2642	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+
22 *Alu*I 2814 (C/A)	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
23 *Dde*I 2820 (G/A)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
24 *Msp*I 3082–3085	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
25 *Hha*I 3784 (T/C)	−	+	+	+	−	+	+	+	+	−	−	−	+	−	−	−	−	−	−	−	−	+	−	−
26 *Alu*I 3849–3852	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
27 *Msp*I 4195–4198	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+
28 *Alu*I 4319 (T/A)	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
29 *Msp*I 4921 (T/C)[a]	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+
30 *Hin*fI 4926 (T/A)[a]	+	+	+	+	+	+	+	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+
31 Allozyme	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	S	F	F	F	F	F
32 indelA −45 +49	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
32 indelB 1111–1241	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
33 indelC 1649–1749	0	+1	0	+1	0	+1	0	0	0	0	−9	0	+1	0	0	0	0	0	0	0	0	0	0	0
34 indelD 2413–2522	0	0	0	−26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35 indelE 2522–2608	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
36 indelF 4236–4485	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	+1	0	0	0	+1	0	0	0	0
37 indelG 4336–4416	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Population: Zimbabwe	1	1	1	2	6	1	3	1	1	4	2	1	1	2	3	1	3	7	1	1	2	1	1	1
Raleigh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arvin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Soda Lake	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Asia/France	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	Frequency				
																					Zim.	Ral.	Arv.	S.L.	Asia/Fr.
+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	0.04	0.02	0.00	0.00	0.00
+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1.00	0.96	1.00	1.00	1.00
+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	+	+	0.00	0.38	0.18	0.29	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	0.00	0.00	0.00	0.00	0.14
−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.98	0.75	0.71	0.94	1.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.04	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	0.00	0.02	0.00	0.00	0.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.96	1.00	1.00	1.00	1.00
−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	1.00	1.00	1.00	0.94	1.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.23	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.23	0.00	0.00	0.00	0.00
−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1.00	0.96	1.00	1.00	1.00
+	−	−	−	−	+	+	+	−	−	−	−	−	−	+	−	−	−	+	−	−	0.98	0.70	0.53	0.41	0.57
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.98	1.00	1.00	1.00	1.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.98	1.00	1.00	1.00	1.00
+	−	−	−	−	+	+	+	+	+	−	−	−	−	+	−	+	+	+	+	+	0.29	0.66	0.47	0.41	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	−	−	−	−	+	+	+	+	+	−	−	−	−	+	−	+	+	+	+	+	0.96	0.66	0.47	0.41	0.00
+	−	+	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.00	0.23	0.29	0.06	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	+	−	+	+	+	+	+	+	+	−	−	−	−	+	+	+	+	+	+	+	0.75	0.40	0.18	0.35	0.00
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.02	0.00	0.00	0.00	0.00
+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	0.06	0.00	0.06	0.00	0.00
−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	1.00	0.98	1.00	1.00	1.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.94	1.00	1.00	1.00	1.00
+	−	−	−	−	+	+	+	+	+	−	−	−	−	+	−	−	−	−	−	−	0.06	0.66	0.47	0.41	0.00
F	F	S	F	F	F	F	VF	F	F	F	S	S	S	F	F	F	F	F	F	F	0.02	0.40	0.18	0.29	0.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.02	0.00	0.00	0.00	0.00
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	0.00	0.00	0.06	0.00	0.00
+1	−1	0	−1	0	+1	+1	+1	+1	+1	0	0	0	0	+1	0	+1	+1	+1	+1	+1					
0	0	−40	0	0	0	0	0	0	0	−40	−40	−40	−40	0	−40	0	−10	0	−10	−10					
−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.00	0.00	0.06	0.00	0.00
0	−1	0	−1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.02	0.00	0.00	0.00	0.00
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
2	1	1	11	0	1	12	0	2	0	0	0	1	19	1	2	0	0	0	0	0					
0	0	0	4	1	0	7	1	0	1	0	0	0	3	0	0	0	0	0	0	0					
0	0	0	1	0	0	9	1	0	0	1	1	0	4	0	0	0	0	0	0	0					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3	1	1					

Nucleotide positions and frequencies of four-cutter mutations and composite haplotypes. Frequencies are for the minus restriction site variant, the Slow electromorph, and the least common size variant. Frequencies are omitted for the size variants with more than two states. A base pair substitution resulting in the loss of a restriction site is localized to the four or five bp sequence indicated. A gain of a site could, in most cases be attributed to a base pair substitution at a single nucleotide site, indicated after the restriction enzyme. The mutation, where known is indicated in parentheses, with the nucleotide in the SCOTT and LUCCHESI (1991) sequence given first. Three sites known to be polymorphic in Zimbabwe are not included because two lines were not scoreable for all restriction enzymes. These sites include *Dde*I 2818-2822, *Taq*I 2092-2095, and *Msp*I 2676-2679, all present once in the total sample of 50 chromosomes. Approximate sizes of insertion/deletion variants not specified in the table are as follows: IndelA, 650-750 bp insertion; indelB, 7 bp deletion; indelE, 5 bp insertion; indelG, 2 bp insertion. The specific designations of lines belonging to given haplotypes are presented in the Appendix. In our previous six-cutter survey of the California populations (BEGUN and AQUADRO 1991), four lines were assigned to the wrong electrophoretic class. The correct phenotype for lines A7, A9, A11, and A13 of BEGUN and AQUADRO (1991) are S, F, VF, and F, respectively. These errors led us to underestimate the level of linkage disequilibrium. None of the conclusions of BEGUN and AQUADRO (1991) are affected.

[a] The locations of these site are ambiguous, and are assigned specific locations for ease of presentation. The position of *Alu*I-29 was estimated from the size of an *Alu*I fragment instead of from DNA sequence data, which were not available for this region. *Hha*I 55 is a gain of a site with respect to the SCOTT and LUCCHESI (1991) sequence. The gain may be located at position 53 or 564. The gain of a *Hae*III site (mutation number 18) may be a G to C change at position 2272. The gain of a *Hin*fI site (mutation number 20) may be a T to C change at 2477. The location of *Msp*I 4921 and *Hin*fI 4926 (mutations 29 and 30) are inferred from our unpublished sequence data from the 3′ flanking region of *Pgd* and should be considered tentative. These ambiguities do not affect any analyses or conclusions.

```
466   ACAAGCGTGTCGAA.CCTTCGTGTACGAGTACACCTCCAAGCTGAAGGGCAGCATTA.GTCGGAGCAGCGGCATTGGCTTCCTGAAGAAGGATTACCTGCACTACTCCAAGGACCCGGTGG
           G     C     C    .                           T G                        C

585   CCATTGGCTACATGCGCGAGATGAAGAAGCTGCTGGACCCCAACAGCATCCTCAATCCCTACAAGGTGCTCAACTGAAGGCTTCTAACTAATAGGTTCGATTTTTTTTTGTTTGTGTATA
                                                                          T          C      A  T .                G

705   ATTTGTATAACCTTATAATACAGAAATGGCTTTAAAACTTAATTTTTTTAAACTTGTGGAGTTATAAAGGACCATCATATATGGCACGAAACCAATGGGCAAAAGCTGCTTATAGAATAG
        TC                       A  G  G G     G .      A     A           T                      CT A    A

823   TTCGAAAAAA....ATAGCGCTACCAGTTTTTAGAGTACCACATTTTAGGT.........TAATACCAATAGGAAAAACACTATCTTTGCATTTACTGTTATACTACGAAGCACTATATT
       C       TAGT   ....          C         A      ACTCGGAGG    ...    A                         C

936   TAGTAGCAGTACCCATTAGAGTCCAAGGACTCACAACACTAGCAACAACCGGCATTAATTGCTTAAGGATCGGCGATAGC..............TTCAAAGTCAAACTGCCGCCCTGCGA
                T          ....       A T     C        G     C C              TTCAAACTCCGAAG

1052  AACGTTCGCGAGTGGAGCTTTTCCGCACTTATCGATAGCTAACATTGTGGCGCGACTATCGATCGACGAGCTGGCGCTTAACAGTGCCATATATAGATTGTAACATTGGGAACTCAAATC
       GC               T                                               C                             A  G

1172  ATTGTTGGAACACAAACCGCACGAAAAACACACAGAACATGAGCGGGTGAGTAGAGGGAAATGCCCTTTTCCCTGGAGTTTTCCCGCGATCCTAACGTCGCCCATCCCCGGATCTCTTCCA
                   ..  A  G      GA                             T T        C                             TT      T

1290  GACAAGCGGACATCGCCCTCATCGGTCTGGCCGTCATGGGCCAAAACCTGATACTTAACATCGACGAGAAGGCGATTCGTGGTGTGCGCCTACAACCGCACGGTGGCTAAGGTCAAGGAGT
           T  T         C                         C                                                     C
                       Gly   Asn                 Lys
1410  TCCTCGCCAATGAGGCCAAAGGCACCAACGTGATTGGCGCCGACTCGCTCAAGGACATGGTCTCCAAGCTAAAGAGCCCCCGGAAGGTCATGCTGCTGGTCAAGGGTGAGTTGCTAATCC
                T  G A      A      A        G                G                                           AT

1530  AAATTCAGCGGCTGGGTAGCGCAGAACATCGAAAACCCATCGAAAACCTGCTGCAAGCGATCGCTGTGTTGGTGACTCAGTTTACATGTGTGCGCGCGTGCTGGTGAATTGGTGAAAAAG
                          G          T .                                    AC                   T

1649  TCAAAGCAAAGTCATCA....AATGATTTTTGCGGCTCATATTGCAATGTGCAAAGGGTAAAGAGAGGATAAGCAGGT.GAAGTAGGAAGAGGAGGGAGTGCGAGAAAACGGTGAAGGGG
       G              TGATG C           C            G C  T        G                    C                 G

1769  TGGAGTAAACACAAAAAACAACTGTTTAGAGATCATTGAGCTGCGGCCTTAACCACTAAGTCTGGGTTAGGAGTTGTTATTGTACACATGATACGGATGGCACTTCGTAGGTTTTGTATG
                                                                   C  T      A                 T

1889  CGATAGTTTTGACTTAAAAACT.......................................AGCTCTTTTAGATCCGCACTTCTGATAAAAGGTTTATAATTGAATCAGTCCCA
                G       AGACAGAATTCAAAGCTCTTTATAGGATTCGAAACTTAAGCATTG           A A                            A  G

2009  TTTCGAAAAACCGTAACCGCAAGTCTGGCGTGAGTTTTATCTAATCACATAGTTGTGGTAAGCTGCCTCCACATACCTAAACCATCGAGCCAACGCATAAGGTGATTTCCAGGTCACTCAC
                    A    G        G                                    T              G   C   C

2129  GCGTTCGTCTATCACACTCGCGTGTCCGAAACTCCGCTCACCTCTAGATCGGCGTGCCCGGCTTATCTCCACCTCCCTACCTAGCTAGCTGTTCGTGCGAAAGCAACAACAACCCGGCGC
              C  T              T                            ......................                          G

2227  AGAGAGAAATCTTTGACATTCATTATAGGTCACACAAAACGGGCGATTTT............CTCGGATTTGACCAGCCGAAAAACTTACAGTTTCCTTCCCTGCGAGTTGCCAGGCGAG
                       A           T        CAGGTGGATTTA                            C   TA    CTT           C

2347  TGAGTCATTTCGTCTGGAGACTGCTCCTTAGAAGAATACAGTGCGGGTCAATAACATATGTACATAGCTCTGGAGGTTTTTGTGCTGAACTTATGTAGATTTGAAAGTTGCGTGACAGGT
                                                                                            A

2467  TTTGCGAACTCCCACTTTCACAGGGTGGGTGTTAGTAAGGACAAAGGTGATTGAATATCAGTGTAGTTAAGTTACCAAGGATGACGATACAAAAAACGCTAGTTGATTATTGAACAGAGC
       G    T    A          GG   ......................................          C       ..      G C

2545  GAGTCCCACAATCTTGACCGCTAGTGTGTCC..CAGACCAGCACCAACGACCGCTAGATAGATAGATCAATGCTATTATCGCCACGACTCGTTGGCCTTATCTGGTTCCACTGCGCTGGA
          A                  C  CA  A   C                         G  G                      G

2665  GAAGTGCTCACCCGGCGCAAGGGACGCTTCACCACCAGGGTGACCTCCAAAACAAAACAGTGACTTTCCAAAAACAACCAACCGACCTTGTCCAATCGCACGCTCTACTTTCTGGCTAAA
        C          T      C  G           .....            G

2780  TCGATGATTTCTTCACGT..AAAGAACTGGCGCTTGCTAAGAACGCTTTGGAAAATGAGTCTACGATAGTTGAGCTTTAGAGAACCCCGATGAGGAATTCGATCTGCTTTAATCGACTTT
        A     T   TA           C                                         T          C

2900  ACGCTAATAAGAAATAAACTCGGTCCAATTTCAGCTGGAAGTGCGGTCGACGACTTCATCCAGCAGCTGGTGCCGCTGCTTTCCGCCGGCGATGTGATCATCGATGGCGGCAACTCGGAA
            C  TG     A A           A                                                       T       G
                                            Tyr
3020  TATCAGGACACATCTCGTCGCTGCGACGAGTTGGCCAAACTTGGCCTGCTCTACGTCGGATCCGGCGTGAGCGGTGGCGAGGAGGGCGCCCGCCATGGACCCTCGCTAATGCCCGGCGGA
                C          A           T                                              C      G

3140  CACGAGGCCGCATGGCCCCTCATCCAGCCCATCTTCCAGGCGATCTGCGCCAAGGCCGACGGTGAACCCTGCTGCGAGTGGGTGGGCGATGGAGGCGCCGGTCACTTCGTCAAGATGGTA
                G     T   A                                                                                G
                                             Gln
3260  CACAACGGCATCGAATATGGTGACATGCAGCTGATCTGCGAGGCGTATCACATCATGCAGAGCCTGGGACTGTCGGCTGACCAGATGGCAGACGAGTTCGGCAAGTGGAACTCGGCCGAA
                C                        C     A
                                             His
3380  CTGGACTCCTTCCTCATCGAAATCACGCGTGATATTCTTAAGTACAAGGACGGCAAAGGTCATCTGCTGGAGCGGATTCGCGATACCGCCGGCCAGAAGGGCACAGGCAAGTGGACGGCC
                T                                            T                                     G      A

3500  ATCGCTGCTCTGCAGTATGGAGTGCCTGTGACGCTAATTGGCGAGGCGGTCTTCTCGCGATGCCTGTCTGCTCTGAAGGCACGAGCGTCTCCAGGCCAGCAGCGTGCTGAAGGGACCCTCG
                                                                    C        C
          Glu
3620  ACCAAGGCCGAAGTGGCCAACCTAACCAAGTTCCTCGACGATATCAAGCACGCTCTCTACTGCGCCAAGATCGTGTCCTACGCCCAGGGATTCATGCTCATGCGAGAGGCGGCCAGGGAG
          GC         C         C
                                                                                                       Gln
3740  AACAAGTGGAGACTCAACTACGGCGGAATTGCGCTGATGTGGCGCGGCGGCTGCATCATCCGCAGCGTCTTTCTGGGCAACATCAAGGACGCGTATACGTCGCAGCCGCAGCTGTCCAAC
               T  T     C          T                         T                         G     T  T

3860  CTGCTGTTGGACGACTTCTTCAAGAAGGCCATTGAGCGCGGCCAGGACTCGTGGCGCGAGCTGGTGGCCAATGCCTTCCGCTGGGGCATTCCCGTGCCGGCCCTGTCCACTGCCCTTAGT
        C   T          C                                                                          T  C     A  C

3980  TTCTACGACGGCTACCGCACGGCCAAGCTGCCAGCCAACTTGCTGCAGGCCCAGAGGGATTACTTCGGCGCCCACACCTATGAGCTGCTGGGCCAGGAGGGTCAGTTCCACCACACCAAC

4100  TGGACGGGCACCGGCGGCAATGTGTCCGCCAGCACCTACCAGGCGTAGGTTCCACCTGCTCTA.TTTTCAGCTCACACATTCCATGTCATTGGCGCCGGTGTCTTAGATGTTTCTTTTTT
       A                       T                          C C   C C T
```

FIGURE 2

TABLE 2

DNA sequence polymorphism in a North Carolina sample of *Pgd* in *D. melanogaster*

	Line number and electromorph class												
	Slow					Fast							
Nucleotide	1	14	19	38	42	2	3	8	9	12	41	44	48
781	A	A	A	A	A	A	C	A	A	A	A	A	A
843	C	C	C	C	C	T	C	C	T	T	T	C	C
927	C	C	C	C	C	T	C	C	T	T	T	C	C
1171	A	A	A	A	A	A	A	G	A	A	A	A	A
1661	C	C	C	C	C	C	C	A	C	C	C	A	A
1692[a]	C	C	C	C	C	-	C	C	-	-	-	C	C
1710[b]	-	-	-	-	-	-	-	AG	-	-	-	AG	AG
1764	G	G	G	G	G	G	G	A	G	G	G	A	A
1955[c]	-	-	-	-	-	-	-	T	-	-	-	T	T
1972	C	C	C	C	C	C	C	G	C	C	C	G	C
2322	C	C	C	C	C	C	C	C	C	C	C	C	T
2476	T	T	T	T	T	T	T	C	T	T	T	C	C
2814	C	C	C	C	C	C	C	A	C	C	C	A	A
2820	G	G	G	G	G	A	G	G	A	A	A	G	G
2917	A	A	A	A	A	A	A	T	A	A	A	T	T
3317[d]	A	A	A	A	A	C	C	C	C	C	C	C	C
3784	T	T	T	T	T	C	C	C	C	C	C	C	C
3871	T	T	T	T	T	C	C	C	C	C	C	C	C
4135	T	T	T	T	T	T	T	C	T	T	T	C	C
4217-24[e]	9T	9T	9T	9T	9T	10T	10T	8T	10T	10T	10T	8T	8T
4466-7[f]	GC	GC	GC	GC	GC	-	GC	GC	-	-	-	GC	GC
4505	C	C	C	C	C	T	T	T	T	T	T	T	T
4506-11[g]	T	T	T	T	T	T	T	-	T	T	T	-	-

The following 40 nucleotides are inserted between nucleotides 2502 and 2503 in lines CAM 2, 8, 9, 12, 41, 44 and 48: AAGGACAAGGTTATTGATTATCAGTGTAGTTAAGTTATC. This 40 nucleotide segment corresponds to indelD from Table 1 and is not present in the sequenced *D. simulans* allele.

[a] Either nucleotide 1692 or 1693 is deleted.

[b] AG inserted between nucleotides 1710 and 1711.

[c] T inserted between nucleotides 1955 and 1956.

[d] Site 3317 accounts for the F/S electrophoretic polymorphism.

[e] Length variation in a run of Ts from 4217–4224 of SCOTT and LUCCHESI (1991).

[f] Deletion of bases 4466 and 4467.

[g] Length variation in a run of Ts from 4506-4511 of SCOTT and LUCCHESI (1991).

quency greater than 0.1. All 55 pairwise comparisons had D' of +1 or −1, however, because of the smaller sample size, only 43 of the comparisons were significant ($P < 0.05$).

Most of the Zimbabwe variants were rare; only sites *Alu*I 1808, *Taq*I 1939, *Hin*fI 2473 and *Hha*I 3784 were included in the linkage disequilibrium analysis. Five of the six pairwise comparisons (*Taq*I 1939 *vs.* *Hha*I 3784 was the exception) were significant ($P < 0.05$), with four of these five comparisons having a D' of +1 or −1. The one significant comparison showing all four gametic types was *Hin*fI 2473 *vs.* *Hha*I 3784. This was also the only pairwise comparison for which the Zimbabwe and U.S. samples could be compared.

```
4220   TTCTGGAGTACTTTAGTACTTATTTATACCATTAATATATATGTATGTATATAGAATTTCATAATTGTTGTTAAACATAACATTAAATTGGTGTTTTTTTTGCTAGCAAATGATTTTGAT

4339   TCCTTGGTTTCATGGATGCAAGTGCCATTTAAAATCAACAATGCGTGTGGTGTGGTGTGTTGTATTGTGTGTGGGTCGAGTCTTTCGAGTTTTGTCTTCATCTGGAGACGCCTCCTGCTC
            A        A          A                         T           G                           G

4459   CTTCTACCGCTCCTTCCCTGCTATTGTACTCTCTTCAGCTAGCGCG..TTTTCCGCTCCTTATTTCCCTCAGTCGTCCGAGGGCTTCAGGGTCTTCTTGTTCTCTATGACCAGTTTGTCA
                                                     CT    T      G         T                                     A

4579   GCGGAGTACAGGTGGCCGATGATTACCTGTGGACATTCAAAGGTTAATAAACTCAACCGGCTGATAAGCG.AAAAGGGGCAAAATGGTTACTTTCGATTTCTAATAGGATGGTAATTGAG
            A                                                                A

4699   TTTTCCATTCCCCATATTTGCAAAATCAGATATATATGATAAAATCTACTTTAAATATACATTAATGTT
                                                                         A
```

FIGURE 2.—DNA sequence of the *Pgd* region from *D. simulans*. Differences between the *D. simulans* and *D. melanogaster* Slow sequence (SCOTT and LUCCHESI 1991) are shown below the *D. simulans* sequence. Replacements substitutions are shown above the sequence. Deleted nucleotides are indicated by dots. Positions of the first nucleotide of each line are indicated and follow the numbering of the *D. melanogaster* sequence of SCOTT and LUCCHESI (1991). The first, second, and third exons include nucleotides 1208–1215, 1291–1514, and 2934–4144, respectively (SCOTT and LUCCHESI 1991).

TABLE 3

DNA polymorphism at the *Pgd* locus in a North Carolina sample of *D. simulans*

		Haplotype		
		1	2	3
a	*AluI* 5′[a]	+	−	+
b	*HhaI* 5′[a]	−	+	+
c	*HhaI* 835–839	+	−	−
d	*HinfI*[b]	+	+	−
e	*Sau3A* 1279–1282	+	−	+
f	*ScrFI* 1315 (T/C)	−	+	+
g	*Sau3A* 1966–1969	+	−	+
h	*TaqI* 2092–2095	+	−	+
i	*MspI* 2220–2223	+	−	−
j	*RsaI* 3257–3260	+	−	−
k	*HhaI* 3586 (T/C)	−	+	+
l	*HaeIII* 3625–3628	+	+	−
m	*AluI* 3979 (T/C)	−	+	+
n	*TaqI* 4896 (T/C)	−	−	+
o	*MspI* 3′[a]	+	−	+
p	indelA 820–980	−	+	+
q	indelB 2811–2879	−	−	+
r	indelC 4424–4485	−	+	−
n		11	7	1

Mutations resulting in a loss of a site are localized to a stretch of 4 bases. Gains of sites are localized to precise nucleotide positions, with the change indicated in parentheses. The first nucleotide in the parentheses refers to the nucleotide found in the sequenced *D. simulans* allele. The location of site n is determined from our unpublished sequence data from the 3′ flanking region and should be considered tentative. All gains of sites are silent changes. Insertion/deletion variants are located between the indicated nucleotides. Site indelA is an approximately 5-bp insertion; indelB is an approximately 2-bp insertion; indelC is an approximately 3-bp insertion. Size variants are insertions relative to the most common state in the sample.

[a] Locations of polymorphic sites a, b, and o are unknown; they are spanned by the probe but are not in a sequenced region. Polymorphisms a, b, and o are as follows: a = loss of an approximately 205-bp fragment and the appearance of an approximately 415-bp fragment; b = loss of a 575-bp fragment and appearance of 120-bp and 455-bp fragments; c = loss of a 290-bp fragment and appearance of a large (>1500 bp) fragment.

[b] The presence of the *HinfI* site in *D. simulans* results from a 4-bp insertion relative to the *D. melanogaster* sequence between nucleotides 959 and 960.

In Zimbabwe the gametic types (*HinfI* 2473/*HhaI* 3784) +/+, +/−, −/+, −/− occurred 5, 29, 7 and 7 times, respectively. In North Carolina these gametic types occurred 18, 0, 14 and 21 times, while in California these gametic types occurred 19, 0, 6 and 9 times. This comparison is further evidence of the disparate histories of *Pgd* from Zimbabwe and America.

In *D. simulans*, the variation among the 19 lines was partitioned among only three haplotypes–two common haplotypes and one singleton haplotype. All 91 pairwise comparisons among the 14 polymorphic sites which occurred more than once had a D' of +1 or −1, and all were significant ($P < 0.05$).

Nucleotide variability in *D. melanogaster*: Table 6 shows the per nucleotide heterozygosities as estimated by π (NEI and LI 1979) and θ (HUDSON 1982) for each population, and for U.S. Fast and Slow electromorphs separately. The North Carolina and California populations have similar amounts of variation which are on the low end of previously reported surveys in this species (AQUADRO 1992). Variability measured over equivalent genomic regions were very similar for the four-cutter and direct sequence data from 13 Pgd genes from North Carolina (θ = 0.0015 and θ = 0.0013 for four-cutters and sequences, respectively). The level of variability among the Slow electromorphs is much lower than that among the Fast electromorphs (Table 6). Estimates of π and θ are 0.0018 and 0.0038, respectively, in Zimbabwe. This population, therefore, has a smaller average pairwise difference but a higher proportion of segregating sites than U.S. populations. The appreciable difference between π and θ in Zimbabwe results from the large proportion of rare variants in this population, which make a negligible contribution to the estimation of π but inflate the estimate of θ.

Nucleotide variability in *D. simulans*: Estimates of per nucleotide heterozygosity at *Pgd* (Table 6) are lower than those from *rosy*, *Adh*, *per* and *vermilion* (AQUADRO, LADO and NOON 1988; BEGUN and AQUADRO 1991; MCDONALD and KRIETMAN 1991; AQUADRO 1992; KLIMAN and HEY 1993; D. BEGUN and C. AQUADRO, unpublished results) but higher than those from *cubitus interruptus Dominant* (BERRY, AJIOKA and KREITMAN 1991) and *achaete* (MARTIN-CAMPOS *et al.* 1992). Consistent with previous surveys of several gene regions among U.S. samples

TABLE 4

Nucleotide differences in the *Pgd* region between *D. melanogaster* and *D. simulans*

	5′ Flanking	Transcript: Introns	Transcript: Exons: Silent	Transcript: Exons: Replacement	3′ Flanking
Nucleotide differences	45	82	45	8	16
Nucleotides compared	713	1420	336	1107	619
Substitutions/site[a]	0.066	0.060	0.148	0.007	0.026

[a] Estimated by the method of JUKES and CANTOR (1969). Consensus intron splice junctions were omitted from the analysis.

TABLE 5

Tests for population subdivision at *Pgd* among *D. melanogaster* samples from California (CA), North Carolina (NC), and Zimbabwe (Zim.)

Samples compared	Probability				
	K_S*	Z*	H_S	X^2	F_{ST}
Arvin, CA *vs.* Soda Lake, CA	0.376	0.454	0.392	0.422	0.00
NC *vs.* CA	0.065	0.043	0.014	0.025	0.04
NC *vs.* CA (no Slows)	0.124	0.076	0.045	0.127	0.02
Zim. *vs.* NC	<0.001	<0.001	<0.001	<0.001	0.25
Zim. *vs.* NC (no Slows)	<0.001	<0.001	<0.001	<0.001	0.35
Zim. *vs.* CA	<0.001	<0.001	<0.001	<0.001	0.24
Zim. *vs.* CA (no Slows)	<0.001	<0.001	<0.001	<0.001	0.38

Probabilities are the proportion of trials (1000 trials) in which the simulated test statistic was more extreme than the observed test statistic (HUDSON, BOOS and KAPLAN 1992). K_S* and Z* take into account the numbers of differences between gametes while H_S and X^2 are haplotype statistics which consider only whether two gametes are the same or different. The X^2 value is with lumping of rare haplotypes so the expected number in each sample is at least one. See HUDSON, BOOS and KAPLAN (1992) for details. F_{ST} was estimated by Equation 3 of HUDSON, SLATKIN, and MADDISON (1992) with weighting proportional to sample size for within population diversity. Sequence length variants were excluded from all analyses of population subdivision.

TABLE 6

Summary statistics of DNA variability in the *Pgd* region of *D. melanogaster* and *D. simulans*

			θ	π
D. melanogaster				
Restriction site data				
Zimbabwe		(n = 48)	0.0038	0.0018
North Carolina	Pooled	(n = 53)	0.0021	0.0024
	Fast	(n = 32)	0.0019	0.0022
	Slow	(n = 21)	0.0006	0.0002
California	Pooled	(n = 34)	0.0018	0.0023
	Fast	(n = 24)	0.0016	0.0017
	Slow	(n = 8)	0.0003	0.0002
Sequence data				
North Carolina	Pooled	(n = 13)	0.0013	0.0016
	Fast	(n = 8)	0.0012	0.0015
	Slow	(n = 5)	0.0000	0.0000
D. simulans				
North Carolina		(n = 19)	0.0037	0.0047

θ for four-cutter data and sequences was calculated according to HUDSON (1982) and WATTERSON (1975), respectively. π was calculated according to NEI and LI (1979).

(AQUADRO, LADO and NOON 1988; BEGUN and AQUADRO 1991; AQUADRO 1992; KLIMAN and HEY 1993; D. BEGUN and C. AQUADRO, unpublished results), *D. simulans* has a higher proportion of polymorphic nucleotides than *D. melanogaster* at *Pgd*.

Statistical tests of neutrality: We used the HUDSON, KREITMAN and AGUADÉ (HKA) test of neutrality (1987) to examine whether the amount of nucleotide variability within *D. melanogaster* and between *D. melanogaster* and *D. simulans* is compatible with a strictly neutral model of molecular evolution. For four-cutter variation, the region used in the comparison spans nucleotides 466 through 4767 following the numbering of SCOTT and LUCCHESI (1991). Restriction sites which fell outside of this region were excluded from the analysis.

For *D. melanogaster*, we compared our *Pgd* data to data from the 5′-flanking region of *Adh* (which is probably evolving according to a model of neutral molecular evolution: KREITMAN and AGUADÉ 1986b; KREITMAN and HUDSON 1991) and to the *period* locus (KLIMAN and HEY 1993). Each comparison has advantages and disadvantages. The 11 5′-flanking *Adh* sequences (KREITMAN and HUDSON 1991) have the advantage of a reasonably large number of bases and individuals surveyed but are from several geographic locations. Genetic structure among *D. melanogaster* populations would violate the assumptions of the HKA test and could contribute to rejection of the neutral model. The restriction site data from 5′-flanking *Adh* are from U.S. samples (KREITMAN and AGUADÉ 1986b) but the small number of bases surveyed relative to the sequences decreases our ability to reject the null hypothesis (*i.e.*, neutral, equilibrium evolution) if it is false. The *Adh* locus has the disadvantage of being autosomal, which forces us to make the questionable assumption of equal effective population sizes of males and females in order to compare *Adh* to *X*-linked *Pgd* (BEGUN and AQUADRO 1991) (this is a conservative assumption if males actually have a smaller population size than females and if a departure from the model results from too little variation at the *X*-linked locus. However, the assumption may decrease the power of the test). The *period* sequences have the advantage of being *X*-linked and show no evidence of a rejection of a neutral model of molecular evolution, but there are the problems of small sample size and possible population structure (KLIMAN and HEY 1993; HEY and KLIMAN 1993). Differentiation between Zimbabwe and America means that Zimbabwe loci should properly be compared to other Zimbabwe loci in HKA tests. This is not an option at the present time, therefore, failure to reject the null hypothesis for Zimbabwe should be viewed as preliminary. TAJIMA's (1989) test revealed no departure from the neutral expectation at *Pgd* (see Table 8).

We used a modified version of the MCDONALD and KREITMAN (1991) test with polymorphism data only from *D. melanogaster*. The numbers of replacment polymorphisms and differences are 1 and 7, respectively, while the number of synonymous polymorphisms and differences are 3 and 45, respectively. This allele configuration is not significantly different from that expected under a neutral, equilibrium

model (Fisher's exact test, $P > 0.50$). Finally, we noticed that seven of eight amino acid differences between species (one of which is the polymorphic within *D. melanogaster*) result in a charge change. We used the method of HUGHES, OTA and NEI (1990) to test the hypothesis that there are more charge differences between species than expected under a null model. The numbers of radical and conservative sites are 439.33 and 653.33, respectively, while the numbers of radical and conservative replacements are seven (or six if the *D. melanogaster* sequence is a Fast electromorph) and one, respectively. The difference in the proportion of radical changes per radical site and the proportion of conservative changes per conservative site is significant ($P < 0.05$) if one uses a normal distribution for hypothesis testing (although it is not entirely clear that this distribution is appropriate) or by the Fisher's exact test ($P < 0.05$) which does not require normality.

For *D. simulans*, we compared polymorphism and divergence at *Pgd* to that observed at the *rosy* locus (AQUADRO, LADO and NOON 1988). Both samples are from the same North Carolina population. For *rosy* we used the polymorphism and divergence estimates from BEGUN and AQUADRO (1991) (original data from AQUADRO, LADO and NOON 1988). There was a significant departure from the neutral expectation (Table 7). The Tajima test statistic for *Pgd* in *D. simulans* was large and positive but not significantly different from zero (Table 8). The distribution of variation in *D. simulans* was reminiscent of that seen among the *D. melanogaster* Fasts in America, with relatively few variants, most of which occurred at intermediate frequency.

DISCUSSION

Variation between species: The overall substitution rate per site at *Pgd* between *D. melanogaster* and *D. simulans* is typical of that seen in comparisons between these species at other genes (LANGE, LANGLEY and STEPHAN 1990; KREITMAN and HUDSON 1991; BERRY, AJIOKA and KREITMAN 1991; MARTIN-CAMPOS *et al.* 1992; KLIMAN and HEY 1993; LANGLEY *et al.* 1993; HEY and KLIMAN 1993). The distribution of differences between species at the 5′-flanking region or large intron *vs.* the 3′-flanking region is significantly heterogeneous at *Pgd* (Fisher's exact test, $P < 0.01$), even though *a priori* we expected sites in all three regions to be under similar selective constraint. The significant excess of charge differences at *Pgd* provides some support for the notion that positive selection is driving amino acid charge substitutions at *Pgd*.

Comparison of four-cutter variation in *D. melanogaster* samples from North Carolina and California: North Carolina and California populations show no significant differentiation for neutral markers at the *Adh* or *period* loci (KREITMAN and AGUADÉ 1986a; BEGUN and AQUADRO 1991; HUDSON, BOOS and KAPLAN 1992). However, North Carolina and California *Pgd* samples are significantly different (*i.e.*, they do not represent two independent samples from a single homogeneous population). The lower frequency of the Slow electromorph on the West Coast is responsible for some of the heterogeneity and is consistent with previously collected allozyme data from *Pgd* (SINGH and RHOMBERG 1987a). If North Carolina and California populations exchange genes frequently and/or have a very recent ancestor, then differentiation at *Pgd* may result from selection on the *Pgd* replacement polymorphism or on a linked site. Furthermore, there is evidence (albeit weak) of geographic heterogeneity at *Pgd* when Slows are eliminated from the analysis. This is because of the different frequencies of the common Fast haplotypes (28 and 31; Table 1). Interpretation of these data may be compromised to an unknown degree by the histories of the lines. The California *X* chromosomes were extracted immediately after the isofemale lines were established while the North Carolina *X* chromosomes were extracted from isofemales lines which had been maintained in the laboratory for three years. Therefore, it is possible that laboratory selection imposed on *Pgd* or a linked site (or sites) may have contributed to differences between California and North Carolina (although it is clear that levels of variability are virtually identical in the California and North Carolina samples). This seems unlikely since selection would have to be very strong to overcome the effects of drift in laboratory cultures with small population sizes. Additional sampling will be necessary to settle the issue.

Comparison of four-cutter variation in *D. melanogaster* samples from Zimbabwe and America: Allozyme studies of *D. melanogaster* suggested that most nuclear gene variation is found within, rather than between populations and that levels of gene flow are roughly equivalent within and between continents (SINGH and RHOMBERG 1987a,b). This is consistent with some DNA studies at the *Adh* locus (KREITMAN 1983; BÉNASSI *et al.* 1993) but not with DNA studies of other loci which provided evidence of restricted gene flow between America and Africa or Europe (EANES *et al.* 1989; MARTIN-CAMPOS *et al.* 1992). Our data support the latter view.

Estimates of F_{ST} between Zimbabwe and North Carolina or California from our four-cutter data are 0.25 and 0.24, respectively. These values are high compared to previous allozyme-based estimates of F_{ST} among global samples of *D. melanogaster* (SINGH and RHOMBERG 1987b) and to previous estimates of differentiation from restriction site data from more geographically restricted samples (LYNCH and CREASE 1990; HUDSON, BOOS and KAPLAN 1992). Many polymorphisms observed in the Zimbabwe sample (in-

TABLE 7

Estimates of polymorphism and divergence used in the HUDSON, KREITMAN and AGUADÉ (1987) test of neutrality in *D. melanogaster* and *D. simulans* and resulting test statistics

				X^2				
				5′ *Adh*				
				Sequence		Four-cutter	Period	
		Bases surveyed	Variable	Global	U.S.	U.S.	Global	U.S.
Polymorphism								
D. melanogaster								
Pgd								
Zimbabwe	$n = 48$	1216 (RE)	19	0.278	0.042	0.070	1.748	0.168
North Carolina	$n = 53$	1252 (RE)	9	2.893	1.750	1.156	5.743**	2.225
	$n = 13$	4203 (Seq.)	17	4.723*	2.862	2.632	9.721***	3.843*
California	$n = 34$	1248 (RE)	8	2.628	1.626	1.101	5.182**	2.081
5′ *Adh*	$n = 11$ (5)[a]	1243 (Seq.)	30 (18)					
	$n = 81$[b]	414 (RE)	9					
period	$n = 6$ (4)[c]	1868 (Seq.)	28 (13)					
D. simulans								
Pgd	$n = 19$	1200 (RE)	10	6.131* (*vs.* rosy)				
rosy[d]	$n = 30$	504 (RE)	28					
Divergence								
Pgd		4203 (Seq.)	196					
5′ *Adh*		1243[a] (Seq.)	77[a]					
		4052[b] (Seq.)	210[b]					
period[c]		1868 (Seq.)	70					
rosy[d]		477 (RE)	19.2					

The effective number of nucleotides surveyed for restriction site data within species at *Pgd* and *rosy*, and between species for *rosy* are estimated by the method of HUDSON (1982) as described in BEGUN and AQUADRO (1991), and are indicated by "RE." Estimates are slightly different between populations and *Pgd* electromorph classes because the number of observed restriction sites differed between samples. For *Pgd* four-cutter data only restriction sites located between nucleotides 466 and 4767 are included in the analysis. For the North Carolina sample of *D. melanogaster* we are unsure whether the polymorphic site designated *Hha*I 55 site is actually located in this position; it could also be located at 53 or 564. For the purposes of the test we assumed the site was located at 55 and therefore, it was not included in HKA tests. The test result is qualitatively the same if we assume the site is at 564. Data gathered by direct DNA sequencing are indicated by "Seq."

[a] Direct DNA sequence data from KREITMAN and HUDSON (1991). U.S. sample size and data shown in parentheses.

[b] Four-cutter data from KREITMAN and AGUADÉ (1986b) as summarized in HUDSON, KREITMAN and AGUADÉ (1987). Divergence is between a randomly selected *D. melanogaster* allele and a *D. sechellia* allele.

[c] KLIMAN and HEY (1993). U.S. sample size and data shown in parentheses. Divergence is between one randomly selected gene from each species (ME NJ1 and SI CA1).

[d] AQUADRO, LADO and NOON (1988).

* $P < 0.05$, ** $P < 0.025$, *** $P < 0.001$.

cluding two intermediate frequency sites, *Alu*I 1808 and *Taq*I 1939) were not seen in U.S. samples and the two geographic regions share no four-cutter haplotypes (Table 1). Importantly, differences between samples cannot be explained in terms of different electromorph frequencies, as high levels of differentiation are still observed if one considers only Fast electromorphs (Tables 1 and 5); the Fast electromorphs observed in the United States are not simply a subset of those observed in Zimbabwe. Pairwise comparisons of two polymorphic restriction sites (*Hin*fI 2473 and *Hha*I 3784) which occur at intermediate frequency in Zimbabwe and America yield linkage disequilibrium coefficients (D') of +1.00 and −0.41, respectively. Thus, the magnitude and direction of linkage disequilibrium for these sites is different in the two geographic regions. Finally, the frequency spectrum of variation is very different in Zimbabwe and America, with Zimbabwe showing greater skewness toward rare variants (Table 8). These data all support the view that the evolutionary histories of *Pgd* in Zimbabwe and America are distinct.

Are differences between Zimababwe and America specific to *Pgd*? Restriction site data from other loci indicate that the answer is no, and that major patterns of differentiation between populations are qualitatively similar for several loci across the *X* chromosome (BEGUN and AQUADRO 1993). This means that at least some of the differences between Zimbabwe and the United States at *Pgd* probably result from general, population phenomena. Higher nucleotide heterozygosities in Zimbabwe (BEGUN and AQUADRO 1993) can be interpreted as evidence that *D. melanogaster* populations in the Zimbabwe region of Africa are histori-

TABLE 8

TAJIMA (1989) test of neutrality on restriction site and DNA sequence variation in the *Pgd* region of *D. melanogaster* and *D. simulans*

			S	k	D
D. melanogaster					
Restriction sites					
Zimbabwe		(n = 48)	22	2.670	−1.497
North Carolina	Pooled	(n = 53)	13	3.755	0.922
	Fast	(n = 32)	10	3.375	1.217
	Slow	(n = 21)	3	0.286	−1.727
California	Pooled	(n = 34)	10	3.526	1.361
	Fast	(n = 24)	8	2.553	0.616
	Slow	(n = 8)	1	—	—
Sequences					
North Carolina	Pooled	(n = 13)	17	6.716	0.954
	Fast	(n = 8)	13	6.357	1.357
	Slow	(n = 5)	0	—	—
D. simulans					
North Carolina		(n = 19)	15	6.351	1.794

S, k, and D, are the number of segregating sites, the site heterozygosity, and the Tajima test statistic, respectively. None of the values are significantly different from zero (the expectation under a neutral, equilibrium model).

cally larger than those from the United States, while the consistent skewness toward rare variants across loci in Zimbabwe (BEGUN and AQUADRO 1993) can be interpreted as evidence of an expanding population (or, less likely, given that many variants are located in introns or flanking sequences, that restriction site variation is deleterious; TAJIMA 1989). The larger decrease in θ compared to π in Zimbabwe and the United States supports the notion of a partial bottleneck along the lineage(s) leading to New World *D. melanogaster* (MARUYAMA and FUERST 1984; DAVID and CAPY 1988; BEGUN and AQUADRO 1993). Theory predicts that dramatic population expansion (as is thought to have occurred in temperate *D. melanogaster* populations) results in an excess of rare polymorphisms (MARUYAMA and FUERST 1984; TAJIMA 1989). This prediction may not be satisfied for U.S. samples because the expansion has been too recent on the time scale of molecular evolution or because our simple ideas about the population dynamics of *D. melanogaster* are incorrect.

Pgd was one of only a few genes showing dramatic electromorph frequency differences between Africa and America in previous allozyme surveys (SINGH, HICKEY and DAVID 1982; SINGH and RHOMBERG 1987a), while our data (BEGUN and AQUADRO 1993; this report) show that significant differentiation is probably genome-wide. There are at least two hypotheses to explain these results. First, the Zimbabwe population may be more isolated from America than other African populations previously examined with allozymes. Therefore, unlike previous contrasts between Africa and America for allozymes, the Zimbabwe sample may show differentiation at silent and replacement sites at many genes when compared to America. African samples showing evidence of differentiation from America are from southern or eastern Africa (EANES *et al.* 1989; BEGUN and AQUADRO 1993) while a western Africa sample showed little evidence of differentiation from U.S. samples at the *Adh* locus (BÉNASSI *et al.* 1993); the allozyme survey of SINGH, HICKEY and DAVID (1982) was also from a western Africa population. Second, it is possible that widespread selection on allozyme variation is leading to the contrast between our DNA data and the previously collected allozyme data. These alternatives can be tested with direct sequence data from several genes from the Zimbabwe sample, other African samples and U.S. samples. Though our data from France and southeast Asia are based on very small samples, they are consistent with some differentiation between these populations and those from America and Africa. Whatever the case may be, the data provide compelling evidence of different evolutionary processes in different *D. melanogaster* populations. Future evolutionary genetic models for this species should incorporate the notion that *D. melanogaster* populations are not at equilibrium for nuclear genes.

Variation in *D. melanogaster* samples from the America: The structure of variation at *Pgd* is unusual compared to that seen in previously examined regions. First, there are very high levels of linkage disequilibrium. We observed more linkage disequilibrium at *Pgd* than was previously observed over a comparable physical distance between *yellow* and *achaete* (MARTIN-CAMPOS *et al.* 1992) even though crossing over at *Pgd* is thought to be higher than at *y-ac* (BEGUN and AQUADRO 1992); in a strictly neutral, equilibrium model one expects linkage disequilibrium to decrease with increased crossing over (though the variance of linkage disequilibrium is probably very large; HUDSON 1993). Second, much of the variation in America is found between two Fast lineages (four-cutter haplotypes 31 and 28; DNA sequences CAM 8 and CAM 9: Tables 1 and 2; Fig. 3). The large four-cutter sample and the smaller direct sequence sample suggest that most intermediates between these two lineages are not present at appreciable frequency in America.

The level of DNA polymorphism among *Pgd* sequences from America is toward the low end of the range reported for previously surveyed regions (reviewed in AQUADRO 1992). Since the *Pgd* region appears to have low rates of crossing over, reduced variation at this locus supports the idea that recombination rates are important determinants of standing levels of DNA polymorphism in natural populations (BEGUN and AQUADRO 1992). Polymorphism and divergence at *Pgd* are incompatible with a neutral, equilibrium model of molecular evolution in some, but not all comparisons to 5′*Adh* and *period* (Table 7).

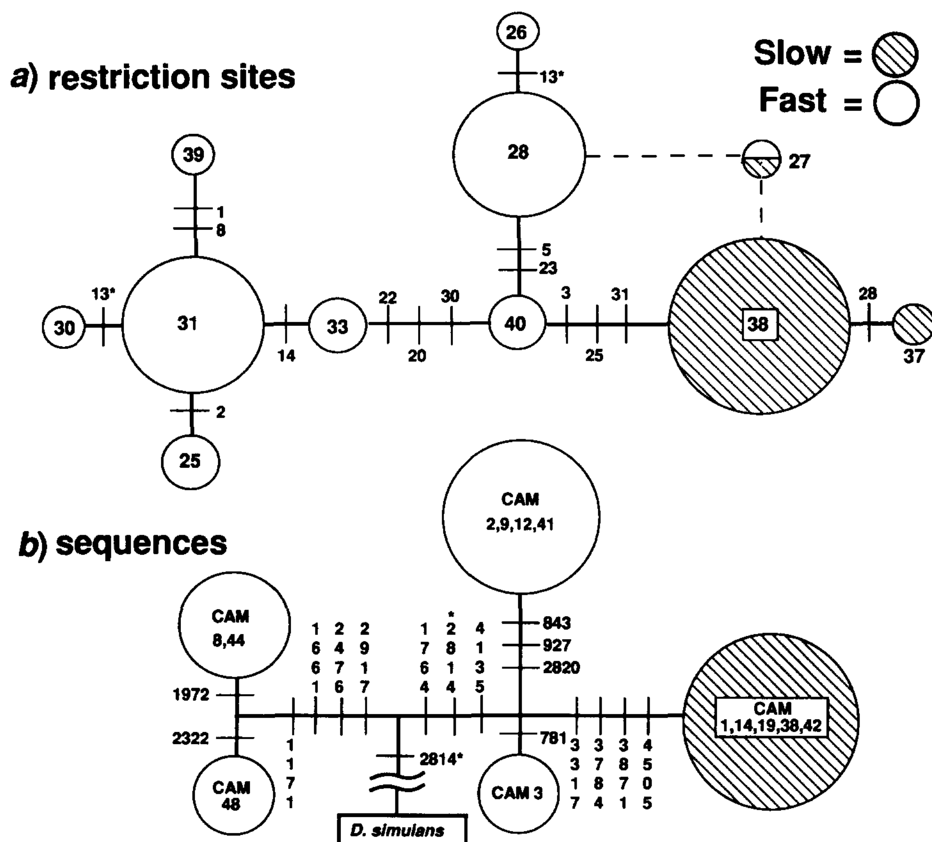

FIGURE 3.—Networks showing hypothesized evolutionary relationships among *Pgd* haplotypes in *D. melanogaster* were constructed using PAUP (SWOFFORD 1991). Surface areas of circles are roughly proportional to haplotype frequencies. Mutations are indicated by tick marks. Sites which change more than once are indicated with an asterisk. (a) Single shortest unrooted network using polymorphic four-cutter restriction sites from a sample of 53 *Pgd* genes from North Carolina. Haplotype and restriction site designations are from Table 1. Haplotype 27 is a putative recombinant between Fast and Slow. (b) Single shortest network among 13 sequnced *Pgd* genes from Norh Carolina using *D. simulans* as an outgroup. Line and nucleotide site desginations are from Table 2. CAM 2, 9, 12 and 41 belong to four-cutter haplotype 28. CAM 1, 14, 19, 38 and 42 belong to haplotype 38. CAM 3 belongs to haplotype 40. CAM 8 and 44 belong to haplotype 31. CAM 48 belongs to haplotype 33.

The heterogeneous test results have several possible explanations. First, the four-cutter data provide less power to reject the null hypothesis as a consequence of the smaller number of bases surveyed. Second, we need to be concerned about the effect of population subdivision. The U.S. *Pgd* sample may have reduced variation because it is from a single geographic location, while the 5′*Adh* and *period* data are from global samples. KREITMAN (1983) has argued that his data support the hypothesis that there is no differentiation among the populations from which his sequences were sampled. We found no evidence of differentiation between the five 5′*Adh* sequences collected in America and the remaining six sequences from Kenya, France and Japan (KREITMAN and HUDSON 1991) using the methods of HUDSON, BOOS and KAPLAN (1992) [$P = 0.95$ for K_S* and $P = 0.90$ for Z*], suggesting that our test of 13 *Pgd* sequences *vs.* 11 5′*Adh* sequences is appropriate. However, given that we may have limited power to detect differentiation with small sample sizes, a conservative test would compare our North Carolina *Pgd* sample to a more geographically restricted sample of 5′*Adh*. Comparison of North Carolina *Pgd* sequences with U.S. 5′*Adh* sequences does not reject the HKA model (Table 7), however, this may simply be a consequence of the reduction in sample size compared to the global 5′*Adh* sample. Comparison of *Pgd* and *period* sequences rejected the HKA model, while comparison of *period* and 5′*Adh* revealed no departure from the HKA model ($X^2 = 0.57$). A final comment on the HKA tests is with regard to the hypothesis of a population bottleneck along the lineage leading to U.S. samples (see above). While the HKA test is probably not seriously compromised by demographic phenomena such as bottlenecks (as long as the loci being compared are from the same population), additional studies of the behavior of the HKA test statistic under various violations of the assumption of equilibrium would be worthwhile. Nonetheless, the weight of the evidence suggests that variation at *Pgd* is inconsistent with neutrality, probably as a consequence of too little variation within U.S. *D. melanogaster* samples.

There is very little DNA variation segregating among the Slow electromorphs, even though they occur at intermediate frequency in America. If we "undo" the putative recombinant in the North Carolina four-cutter sample there is only one polymorphic site among the 21 Slows and there are no segregating sites among 4203 bases surveyed in a sequence analysis of five Slow genes from North Carolina. To explore whether the Slow electromorphs have significantly less variation than expected, we calculated the expected number of segregating sites in the Slows following HUDSON and KAPLAN (1986). Using their notation for our four-cutter data, $k_0 = k_1 = 21$, $n = 53$, and the

per locus estimate of θ_2 from four-cutter data is 3.3. Then, the expected total time in the genealogy of the Slows (in units of 2N generations) from their equation 16 is 2.8 and the expected number of segregating sites is 4.7. Given the large variance associated with the expectation, 4.7 is probably not significantly different from the observed value of one. It is unclear, however, whether (observed-expected)^2/variance is chi-square distributed in a nested subsample (R. HUDSON, personal communication), so hypothesis testing required the following computer simulation.

Neutral, no recombination genealogies of 53 gametes were simulated using the approach of HUDSON (1990). The Slow *Pgd* electromorph is the derived state with respect to the outgroup, *D. simulans* (which is monomorphic at the allozyme level). Therefore, we simulated 5000 genealogies and selected the 1162 genealogies which had a monophyletic group of 21 gametes. The mean and variance of the time in the genealogies for 21 gametes were 1.8 and 0.7, respectively, while the mean and variance for segregating sites were 3.0 and 4.8, respectively. The probability of observing one or zero segregating sites was 0.26. A similar test was carried out on the direct sequence data. The expected number of polymorphic sites among the five Slow sequences determined from simulations was 2.4 while the variance was 5.8; the probability of observing zero polymorphic sites in the Slows was 0.22. Therefore, there is no statistical evidence for significantly reduced variation within the Slow electromorphs. The mean time from simulations was smaller than the time determined analytically because the simulations included the information that Slow is the derived state while the theory developed in HUDSON and KAPLAN (1986) assumes that there is no information on which class is derived. Therefore, the times in the genealogies calculated by the method of HUDSON and KAPLAN (1986) are means of the cases where the class of interest is derived or ancestral, and may overestimate the expected time in the genealogy of a class which is known to be derived. The simulation results are in complete agreement with the analytical results when the simulations allow Slow to be either derived or ancestral.

Our ability to test for reduced variability within the Slows may be hampered by the overall low level of variability in the *Pgd* region. More importantly, since there is some question whether the *Pgd* region is at mutation-drift equilibrium, one must be concerned about the reliability of the estimate of θ from this region. Future tests of the neutral theory might be improved if nested subsamples could be incorporated into an HKA framework. Studies of the distribution of segregating sites in such samples will be required.

While there is no statistical evidence for reduced variability within the Slow electromorphs, overall reduced variation at *Pgd* still demands an explanation. There are several possibilities. Reduced variation could result from the hitchhiking effect of an advantageous mutation at *Pgd* or a linked site (MAYNARD SMITH and HAIGH 1974; KAPLAN, HUDSON and LANGLEY 1989). The absence of a skewed frequency spectrum in U.S. *Pgd* samples appears to be inconsistent with a simple "catastrophic" sweep hypothesis, which would posit that *Pgd* was monomorphic at the fixation time of the advantageous mutant (HUDSON 1990). Perhaps a more realistic model to explain the amount and distribution of variation at *Pgd* under a hitchhiking hypothesis would posit some crossing over between *Pgd* and a selected locus during the sweep (KAPLAN, HUDSON and LANGLEY 1989). Such partial sweeps may be analogous to partial bottlenecks, causing a loss of rare variants (CHARLESWORTH, MORGAN and CHARLESWORTH 1993). While the previously discussed models assume that selected mutants are initially unique, partial sweeps could also be a consequence of selection on low-frequency neutral or slightly deleterious variants which become positively selected subsequent to environmental change. Under these conditions, it may be possible for strong selection with no crossing over to result in a frequency spectrum which is not obviously skewed toward an excess of rare sites. Quantitative treatments of the distribution of effects of hitchhiking vis-à-vis the frequency spectrum will be required to explore these possibilities. Furthermore, though we attempted to rule out bottlenecks in lineages leading to U.S. populations as a sufficient explanation for reduced variation at *Pgd*, such a phenomenon could still have an important effect on the frequency spectrum (BEGUN and AQUADRO 1993). Two additional explanations are that the Slow allele (or a linked site) is currently sweeping through the population in America and will eventually fix, or that the Fast/Slow polymorphism (or a linked site) is a relatively recent balanced polymorphism. Finally, we can not rule out the possibility that background selection against deleterious mutants makes a significant contribution to reduced variability (CHARLESWORTH, MORGAN and CHARLESWORTH 1993).

If levels of variability prove to be low over vast physical distances in the distal part of the *X* chromosome, then it is certainly reasonable to posit that nothing unusual is occurring at *Pgd* and that variability is reduced by some combination of partial sweeps and background selection at linked sites. However, data from biochemical studies, artificial selection experiments, and population surveys (BIJLSMA and VAN DELDEN 1977; BIJLSMA and VAN DER MEULEN-BRUIJNS 1979; BIJLSMA and KERVER 1980; CAVENER and CLEGG 1981a,b; CAVENER 1983; OAKESHOTT *et al.* 1983; SINGH and RHOMBERG 1987a) certainly suggest that the *Pgd* replacement polymorphism or polymor-

phism at a tightly linked site is *currently* under selection as a new balanced polymorphism or an advantageous mutant which is yet to reach fixation. Finally, we note that for *Adh*, *G-6pd*, and *Pgd*, one of the two major electrophoretic types appears at low frequency in Africa, at intermediate frequency in America, and has a low level of DNA variation (KREITMAN 1983; AQUADRO *et al.* 1986; EANES *et al.* 1989; MIYASHITA 1990). This pattern supports the idea that temperate populations are responding to a novel environment with rapid allele frequency changes at several genes (DAVID and CAPY 1988; SINGH 1989). While this might appear to be in conflict with our earlier hypotheses regarding genome-wide differences between Zimbabwe and America, the reality may be that simple explanations for the data are inadequate. That is to say, we should remain open to the idea that the distribution of silent and replacement polymorphism at different genes and in different populations has multiple causes (*e.g.*, KREITMAN and HUDSON 1991).

Variation in *D. simulans*: Four-cutter data from *D. simulans* reveal high levels of linkage disequilibrium and low levels of nucleotide heterozygosity at *Pgd* compared to previously surveyed regions in areas of normal recombination in this species (AQUADRO, LADO and NOON 1988; AQUADRO 1992). Polymorphism and divergence at *Pgd vs. rosy* show a significant departure from the neutral expectation. The cause of the departure is probably a reduction in variability at *Pgd* in *D. simulans*. It will be important to determine whether the reduced variability we observed in *D. simulans* is restricted to the *Pgd* region or extends over a large fraction of the distal portion of the *X* chromosome as previous results have suggested (BEGUN and AQUADRO 1991; MARTIN-CAMPOS *et al.* 1992; AQUADRO and BEGUN 1993). As noted earlier, a simple sweep model incorporating strong selection and no recombination would predict an excess of rare sites. The large positive Tajima test statistic at *Pgd* in *D. simulans* would seem to rule out such a model. Essentially, the same range of possible explanations for the *D. melanogaster* data apply to *D. simulans*.

A final observation on the distribution of variation at *Pgd* in *D. simulans* is a comparison of the present four-cutter results to our earlier six-cutter results (BEGUN and AQUADRO 1991). The large difference in Tajima test statistics from the the two samples ($D = +1.794$ in North Carolina; $D = -1.119$ in California) suggests that the frequency distributions in the samples are different. The samples differ in the geographic location of the collections, the restriction enzymes used, and the size of the genomic region surveyed (a much larger region including flanking sequences were included in the six-cutter survey), any of which may be contributing to the heterogeneity. This apparent difference in the distribution of variation in the samples is worthy of further investigation.

CONCLUSIONS

The Zimbabwe sample of *D. melanogaster* differs from U.S. samples in several respects, including the frequency of allozyme variants, the distribution of haplotypes, the proportion of segregating sites, and the frequency spectrum of variation, all of which suggest that Zimbabwe and America *Pgd* genes have very different evolutionary histories. The *D. melanogaster Pgd* samples from North Carolina and California show high levels of linkage disequilibrium and relatively low levels of nucleotide variability which are probably not attributable solely to partial bottlenecks along the lineages leading to U.S. populations. Thus, we suspect that the distribution of variation at *Pgd* in Zimbabwe and America is the consequence of a complex combination of the histories of the populations as well as selection at *Pgd* or a linked site. *Pgd* and *Adh* show remarkably *similar* patterns of allozyme variation (*Adh* data reviewed in VAN DELDEN 1982). However, the very different distributions of DNA polymorphism at these two genes (KREITMAN and AGUADÉ 1986a,b; SIMMONS *et al.* 1989; KREITMAN and HUDSON 1991) suggest that similarities at the allozyme level may be misleading and that allozyme-coding genes showing similar patterns of phenotypic variation may have very different evolutionary histories. There is significantly reduced variability at *Pgd* in *D. simulans*. The frequency distribution of the variation is unusual, with most variants at intermediate frequency. These two observations considered jointly suggest that the evolutionary history of this genomic region in *D. simulans* is complex. Hopefully, additional sequence data will further illuminate the history of *Pgd* in Africa, North America, and other geographic regions in both *D. melanogaster* and *D. simulans*.

We thank M. SCOTT and J. LUCCHESI for kindly providing clones and unpublished sequence data, R. R. RAMEY, L. BROWN and the Sengwa Wildlife Research Institute for the Zimbabwe *D. melanogaster* sample, A. BERRY for useful advice on running four-cutter gels, M. FORD for computer simulations, A. CLARK, R. HUDSON, C. LANGLEY, R. LEWONTIN, R. MACINTYRE, anonymous reviewers, and members of the Aquadro laboratory for discussion and/or comments, and I. MONTELEONE for making fly media. This work was supported by a grant from the National Institutes of Health (GM 36431) to C.F.A., a National Science Foundation Doctoral Dissertation Improvement Grant (BSR-9112480) to D.J.B. and C.F.A, and a National Institutes of Health predoctoral traineeship to D.J.B.

LITERATURE CITED

AQUADRO, C. F., 1992 Why is the genome variable? Insights from *Drosophila*. Trends in Genet. **8:** 355–362.

AQUADRO, C. F., and D. J. BEGUN, 1993 Evidence for and implications of genetic hitchhiking in the Drosophila genome, in

Mechanisms of Molecular Evolution, edited by N. TAKAHATA and A. CLARK. Japan Scientific Societies Press, Tokyo.

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction-map variation and divergence. Genetics **119:** 875–888.

AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the *alcohol dehydrogenase* gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the *X* chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129:** 1147–1158.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

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.

BÉNASSI, V., S. AULARD, S. MAZEAU and M. VEUILLE, 1993 Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BIJLSMA, R., and J. W. M. KERVER, 1980 The effect of DDT on the polymorphism at the *G6pd* and *Pgd* loci in *Drosophila melanogaster*. Genetics **103:** 447–464.

BIJLSMA, R., and W. VAN DELDEN, 1977 Polymorphism at the G6PD and 6PGD loci in *Drosophila melanogaster*. I. Evidence for selection in experimental populations. Genet. Res. **30:** 221–236.

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.

BINGHAM, P. M., R. LEVIS and G. R. RUBIN, 1981 Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

CAVENER, D. R., 1983 The response of enzyme polymorphism to developmental rate selection in *Drosophila melanogaster*. Genetics **105:** 105–113.

CAVENER, D. R., and M. T. CLEGG, 1981a Temporal stability of allozyme frequencies in a natural population of *Drosophila melanogaster*. Genetics **98:** 613–623.

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

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

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

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

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6.

HARRIS, H., and D. A. HOPKINSON, 1976 *Handbook of Allozyme Electrophoresis in Human Genetics*. North Holland Publishing Co., Amsterdam.

HEY, J., and R. M. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation of multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

HUDSON, R. R., 1982 Estimating genetic variability with restriction endonucleases. Genetics **100:** 711–719.

HUDSON, R. R., 1990 Gene geneologies and the coalescent process. Oxf. Surv. Evol. Biol. **7**.

HUDSON, R. R., 1993 The how and why of generating gene genealogies, in *Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A. CLARK. Japan Scientific Societies Press, Tokyo.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol **9:** 138–151.

HUDSON, R. R., and N. L. KAPLAN, 1986 On the divergence of alleles in nested subsamples. Genetics **113:** 1057–1076.

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

HUGHES, A. L., T. OTA and M. NEI, 1990 Positive Darwinian selection promotes charge diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. **7:** 515–524.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

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

KREITMAN, M., 1983 Nucleotide polymorphism at the *Adh* locus in *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and M. AGUADÉ, 1986a Genetic uniformity in two populations of *Drosophila melanogaster* revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA. **83:** 3562–3566.

KREITMAN, M., and M. AGUADÉ, 1986b Excess polymorphism at the *Adh* locus in *D. melanogaster*. Genetics **114:** 93–110.

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

LANGE, B. W., C. H. LANGLEY and W. STEPHAN, 1990 Molecular evolution of Drosophila metallothionein genes. Genetics **126:** 921–932.

LANGLEY, C. H., J. MACDONALD, N. MIYASHITA and M. AGUADÉ, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90:** 1800–1803.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

LYNCH, M., and T. J. CREASE, 1990 The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol **7:** 377–394.

MARTIN-CAMPOS, J. M., J. M. CAMERON, N. MIYASHITA and M. AGUADÉ, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816.

MARUYAMA, T., and P. A. FUERST, 1984 Population bottlenecks and nonequilibrium models in population genetics. I. Allele

numbers when populations evolve from zero variability. Genetics **108:** 745–763.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

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

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

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

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:** 67–82.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SCOTT, M. J., and J. C. LUCCHESI, 1991 Structure and expression of the *Drosophila melanogaster* gene encoding 6-phosphogluconate dehydrogenase. Gene **109:** 177–183.

SIMMONS, G. M., M. E. KREITMAN, W. F. QUATTLEBAUM and N. MIYASHITA, 1989 Molecular analysis of alcohol dehydrogenase along a cline in *Drosophila melanogaster*. I. Maine, North Carolina, and Florida. Evolution **43:** 393–409.

SINGH, R. S., 1989 Population genetics and evolution of species related to *Drosophila melanogaster*. Annu. Rev. Genet. **23:** 425–453.

SINGH, R. S., D. A. HICKEY and J. R. DAVID, 1982 Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **101:** 235–256

SINGH, R. S., and L. R. RHOMBERG, 1987a A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. Genetics **115:** 313–322.

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

SWOFFORD, D. L., 1991 PAUP (phylogenetic analysis using parsimony) Version 3.0s. Illinois Natural History Survey, Champaign, Ill.

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

VAN DELDEN, W., 1982 The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*. Evol. Biol. **15:** 187–222.

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

Communicating editor: A. G. CLARK

APPENDIX

Iso-chromosomal lines belonging to haplotypes in Tables 1 and 3 are given in Table 9.

TABLE 9

Iso-chromosomal lines belonging to haplotypes in Tables 1 and 3

D. melanogaster (Z, CAM, Arv, SL, Ville, Viet, and Tai refer to Zimbabwe, Raleigh, Arvin, Soda Lake, Villeurbanne (France), Vietnam, and Taiwan, respectively) haplotype

Haplotype	Lines
1:	Z 55
2:	Z 35
3:	Z 47
4:	Z 13, 25
5:	Z 15, 18, 24, 32, 49, 53
6:	Z 6
7:	Z 10, 16, 36
8:	Z 3
9:	Z 48
10:	Z 7, 9
11:	Z 22, 26, 28, 57
12:	Z 41
13:	Z 11
14:	Z 44, 58
15:	Z 31, 40, 56
16:	Z 59
17:	Z 29, 30, 50
18:	Z 5, 12, 17, 21, 23, 42, 43
19:	Z 46
20:	Z 60
21:	Z 27, 39
22:	Z 45
23:	Z 33
24:	Z 34
25:	CAM 50, 52
26:	CAM 18
27:	CAM 28
28:	CAM 2, 9, 12, 23, 33, 34, 41, 55, 61, 69, 71
	Arv 3, 6, 8, 15
	SL 2
29:	Arv 13
30:	CAM 16
31:	CAM 8, 10, 15, 20, 26, 27, 30, 44, 45, 46, 57, 59
	Arv 2, 4, 12, 16, 17, 20, 21
	SL 5, 7, 9, 10, 12, 13, 14, 18, 19
32:	Arv 11
	SL 15
33:	CAM 21, 48
34:	Arv 1
35:	SL 6
36:	SL 16
37:	CAM 54
38:	CAM 1, 5, 14, 19, 22, 25, 29, 35, 37, 38, 42, 47, 53, 56, 62, 64, 65, 66, 68
	Arv 7, 18, 19
	SL 3, 8, 11, 17
39:	CAM 32
40:	CAM 3, 24
41:	Tai 2
42:	Tai 17
43:	Ville 4, 14, 22–24
44:	Tai 9
45:	Viet 15-1

D. simulans haplotype

Haplotype	Lines
1:	CAS 3BX, 10AX, 11AX, 18X, 27AX, 29AX, 41AX, 44AX, 46AX, 53BX, 62BX
2:	CAS 14m, 15BX, 15m, 28AX, 35m, 52AX, 52m
3:	CAS 7m